

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Accurate Antifreeze Protein Prediction Using a LightGBM-Based Model

Rahu Sikander*

Department of Computer Science and Software Engineering, Jinnah University for Women, Karachi, Pakistan Email: rahu.sikander@juw.edu.pk & garahu@sau.edu.pk

Ali Ghulam

Information Technology Center, Sindh Agriculture University Tandojam, Sindh Pakistan

Mujeebu Rehman

School of Information and Communication Engineering, Guilin University of Electronic Technology, Guilin 541004, China

Hina Shafi

Information Technology Center, Sindh Agriculture University Tandojam, Sindh Pakistan

Dostdar Hussain

Department of Computer Science Karakoram International University Gilgit, Pakistan

Syed Saqib Hussain

Department of Computer Science Karakoram International University Gilgit, Pakistan

ABSTRACT

Antifreeze proteins (AFPs) play a role in a large number of organisms such as plants, insects, and fish, in which they prevent ice crystal formation and allow them to survive in very cold temperatures. The AFPs can be used in the metabolic engineering, food engineering, yogurt and cryopreservation. Despite the fact that a number of computational tools have been designed to detect AFPs, most of them are ineffective in prediction with the need to have a more precise prediction technique. During this research, we have reviewed the existing predictors of AFP and we have compared the various machine learning models. We find that LightGBM performs better with pentamer-based feature vectors, which is better than other models on benchmark and independent data, with a sensitivity of 97.50, recall of 98.30, and an F1-score of 99.10. We also emphasize effective classifiers that may be used to enhance the generation of new generation predictors to identify fast and reliable AFPs.

Keywords: AFPs, LightGBM, Antifreeze, Peptides, Bioinformatics, Health Informatics

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

INTRODUCTION

This paper gives a description of the background, biological importance, properties, and uses of antifreeze proteins (AFPs). It also provides a summary of the already developed computational methods of AFP prediction along with their strengths, limitations and opportunities of improvement. The materials and methods section provides information on the datasets employed, the feature extraction strategies employed and the machine learning classifiers employed in the work. We compare the performance of the already existing models, their pros and cons, and define those predictors which allow reaching the best accuracy in AFP classification. This review serves as a good source of information to scholars interested in identifying the right computational tools to use in the process of AFP identification. Moreover, we explore better feature descriptors, feature selection methods, and learning algorithms which could help in coming up with more precise AFP prediction systems.

Antifreeze proteins (AFPs) are present in a variety of living organisms including microorganisms, fish and in plants that live in very cold conditions. Through recrystallization of ice, small ice crystals interplay with the available water molecules resulting in the development of bigger crystals of ice that is capable of destroying the structures of the cells. The AFPs attach to these ice crystals as they are growing and prevent the recrystallization process so the cold-adapted organisms can survive even with the presence of ice under freezing conditions [1]. Among the major properties of AFPs is the reduction in the freezing point of water without altering its melting point, which is referred to as thermal hysteresis. Cell membranes by the lower ectoderm cells are also involved with AFPs in mammalian systems, which protect low temperature injury, but also assist in the maintenance of normal cell biological processes of low temperatures. Other than their biological use, AFPs have been used in cryosurgery, cryopreservation, food preservation and in production of yogurts [2].

AFPs have a number of important properties that have a direct impact on their biological functions. Among them, the thermal hysteresis (TH), affinity of ice planes, and ice recrystallization inhibition are regarded as the most important ones. These properties and the implications of their functional meaning are presented in the subsequent subsections. The difference between the melting and freezing point of a solution is called thermal hysteresis. With this process, AFPs allow marine microalgae, fungi, bacteria, insects, and plants to prevent freezing under colder environments [3]. This property is as well crucial to freeze avoidance in insects and fish, though TH activity in insects is commonly greater than that of purified AFPs and less than the TH activity of fish. The values of TH in the majority of cases are on the rise regarding higher AFP concentration and given an effective range.

Small ice nuclei in the subzero conditions mix with the water molecules and cause recrystallization of ice resulting in the creation of bigger and possibly harmful ice crystals. The AFPs are attached to these developing crystals and prevent recrystallization enabling cold-blooded organisms to live in extreme cold [4]. The other major characteristic of AFPs is that it does not decrease the melting point of water but instead depresses its freezing point. Mammalian cell membranes also deal with AFPs, which shield the cell structures against cold-related damage as well as preserve the

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

physiological function of the cells at low temperatures. They find their use in cryosurgery, cryopreservation, preservation in foods and yogurt making [5].

In freeze-tolerant creatures, the crystals of ice are formed in the extracellular space. Ice nucleation below the freezing point can take place via centipedes, plants, and nematodes using extracellular ice crystals. The bigger ice crystals have lower energy and are thermodynamically stable in comparison to small crystals [6]. Ice recrystallization is the melting of small crystals and growth of the large ones by the Ostwald ripening process. Due to the possibility of large ice crystals to damage cell membrane and adjacent tissues, AFP-positive recrystallization inhibition is a vital prerequisite of cold survival.

Depending on the structural classification, AFPs alter the shape and the development of the ice crystals. The fish AFPs are generally categorized into type I-IV and antifreeze glycoproteins (AFGPs). The Type I AFPs are α -helical and contain a good quantity of alanine, and affect the pyramidal plane of ice. Type II AFPs are globular in nature, and attach to the planes of pyramidal ice. Type III AFPs attach onto both pyramidal and prism planes and type IV AFPs of longhorn sculpin samples exhibit no evident ice-plane selectivity [7].

Significant efforts have been put into the discovery of new AFPs, assessing AFP activity, purification of AFPs, and biotechnological use of AFPs. Purification is usually done through ice-affinity technique whereby AFPs are attached to a cold finger and then pulled out of solution. The methods used to purify the AFPs of cold-acclimated barley include chromatographic methods like ammonium sulfate precipitation, ultrafiltration and MALDI-TOF-MS [8]. Ice-affinity chromatography was also used to purify AFPs of shrub seabuckthorn berries and leaves, and structural characterize it by X-ray crystallography, Raman spectroscopy, circular dichroism, and terahertz spectroscopy [9].

The analysis of AFPs by experiment is frequently expensive and time consuming. Recent bioinformatics and high-throughput technologies are able to investigate AFPs faster and more precisely. The methods have unveiled genome evolutionary signals, identified molecular processes and led to the identification of new AFPs [10]. Transcriptomic studies have provided both functional and evolutionary information about AFP IV, whereas genome pyrosequencing has shown a type of cold-adaptation in psychrophilic yeast. The identification of the gene dosage effects of AFGPs has also been done using whole-genome sequencing. The genetic algorithms, support vectors, Markov state models, and network based methods are other machine learning techniques that have improved AFP prediction and structural analysis [11]. It has created a couple of computational tools, stand-alone software packages, and web servers that have been used in AFP prediction, such as AFP-Pred, *afpCOOL*, Target Freeze, and *iAFP-Ense* [12]. Such tools, combined with state-of-the-art material design including quasi-carbon nitride quantum dots and graphene oxide films, have increased the AFP uses in agriculture, cryomedicine, food processing, and anti-icing technologies.

Two major domains in which AFPs have demonstrated a fair deal of promise are

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

cryopreservation and cryosurgery. The low-toxicity cryoprotective agents are the AFPs which decrease the freezing point without affecting the melting point and stabilize the cell membrane on exposure to cold [13]. Type III AFPs maximise the cryopreservation of the mammalian cells and glycerol is added to maximise the cryopreservation of yeast and red blood cells. More uses of AFPs include fertility preservation, tumor cryosurgery and antibacterial treatment [14, 15]. AFP-transgenic organisms such as fish, fruit flies and crops, among others, have also been shown to be more resistant to frost in agriculture [16]. Tomato seedlings will sprout better when subjected to type I AFPs and transgenic plants of wheat and potatoes endowed with AFPs will improve freeze tolerance at low temperatures [17]. AFPs are useful in preserving food. The type I AFPs decrease drip loss and frozen meat protein degradation. Lamb meat has been injected with AFGPs to inhibit freeze damage. Anti-icing materials that use AFP-like materials are used as anti-icing coatings to enhance visibility on the glass surfaces as is the case with vehicles windshields [18, 19]. Historically, AFPs were initially found in fish of the Arctic [20] and then in fungi, bacteria, plants and insects [21] as well. The AFPs are found in a large range of tissues such as, flowers, leaves, stems, seeds, liver and stomach, and hence their extensive biological importance.

Though they are important, the mechanisms of interaction between AFPs and ice are not fully understood. AFP concentration and exposure time affect the level of activity in thermo hysteresis. The Gibbs-Thomson effect predicts that the farther the molecules of AFP are apart, the higher the TH, and the closer the molecules come to each other [22]. This relationship was experimentally used to measure the surface density of AFPs by Drori et al. [23]. The cryopreservation of oocytes and ovarian tissues is especially difficult because of cryoinjury that is associated with effects on organelles like lysosomes, the endoplasmic reticulum and the spindle apparatus [24]. The AFPs stabilize membranes, maintain the integrity of spindles and enhance developmental competence of oocytes and embryos during cryopreservation [25]. Due to the biological and industrial significance of AFPs, a number of computational predictors have been created. Older strategies, like AFPs [26], used BFDPC, PACAAC descriptors, genetic algorithms and Xgboost, Adaboost, and classifiers. Later approaches embraced physicochemical characteristics, evolutionary profiles, pseudo-amino acid composition, as well as feature engineering mixed with machine learning techniques [27]. The latter models, including LightGBM-based and block-split PSSM-based ones, have demonstrated better performance in terms of accuracy and generalization [28].

Materials and Methods

Data Collection

Both reviewed (Swiss-Prot) and unreviewed (TrEMBL) were taken into consideration to have a good representation of AFPs in various species, including plants, microorganisms, and fish. Relevant keywords like antifreeze protein were searched and filtered by the sequences where the activity is experimentally proven. The critical analysis of the current models is thus very necessary. In this paper, we present summaries of applied datasets, feature encoding schemes, classifiers, and model assessment schemes. We discuss their advantages and disadvantages, the most

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

effective methods, and future perspectives, with respect to feature descriptors and learning algorithms, which can advance future AFP prediction. Table 1 and Figure 1 present a detailed overview of the existing methods, such as the features that were used, classifiers, feature selection methods, as well as evaluation metrics.

Table 1. Overview of Dataset Composition

Datasets	Total	Similarity 90%	Train
Antifreeze proteins	877	481	481
Non-Antifreeze proteins	11493	9493	9493

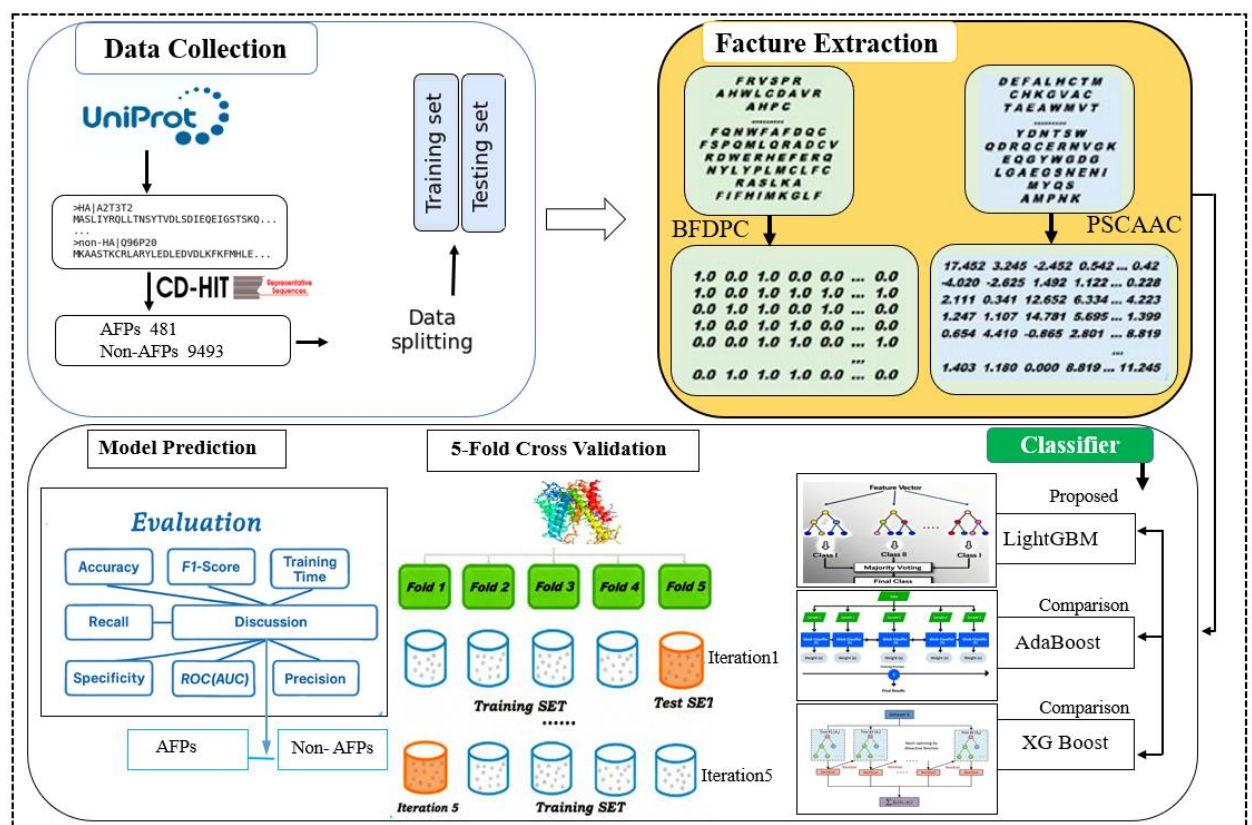


Figure 1. Workflow of the Proposed Model

Feature Extraction

Two feature descriptors were used to numerically encode the antifreeze protein (AFP) sequences to do the computational analysis, i.e. Bi-Functional Dipeptide Composition (BFDPC) and Pseudo-Sequence Composition Amino Acid Composition (PSCAAC). These descriptors describe the compositional as well as the physicochemical

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

characteristics of the AFPs and allow differentiating between functional and non-functional proteins.

Sequence Preparation

Cd-HIT was put in place to eliminate redundant sequences with a sequence identity threshold of 90 to eliminate bias occasioned by very similar sequences. FASTA formatting of protein sequences was then done to analyze downstream. Complete amino acid sequences that have functional annotations and complete sequences were retained only.

Bi-Functional Dipeptide Composition (BFDPC)

The BFDPC records the frequency of all the dipeptides (combinations of two consecutive amino acids) and their physicochemical properties.

Take a protein sequence to be of the form $P=p_1 p_2 p_3 \dots p_L$,

The dipeptide composition is calculated as:

$$D_{ij} = \frac{N_{ij}}{L-1} \quad (1)$$

Where:

D_{ij} is the normalized frequency of the dipeptide composed of amino acid i and j ($i, j \in [?]$) under the conditions of amino acid i and j ($i, j \in [?]$) in A, R, N, D, ... V. In the case of bi-functional aspect, the physicochemical properties like hydrophobicity (H), polarity (P), and molecular volume (V_m) are used to weight each dipeptide producing a feature vector:

$$BFDPC_{ij} = D_{ij} \cdot w_{ij}, w_{ij} = \frac{H_i + H_j + P_i + P_j + V_{m_i} + V_{m_j}}{6} \quad (2)$$

It is a method which considers the order sequence information of the amino-acids and the physicochemical relevance of the amino-acids and it improves the interpretation of the model used to predict the AFP.

Pseudo-Sequence Composition Amino Acid Composition (PSCAAC)

PSCAAC is the extension of the established amino acid composition (AAC), which adds sequence-order data using correlation factors based on physicochemical properties. Where P is a protein sequence, the standard AAC is defined as:

$$AAC_i = \frac{n_i}{L}, i = 1, 2, \dots, 20 \quad (3)$$

Where:

- n_i is the number of occurrences of amino acid i in the sequence,
- L is the sequence length.

To include sequence-order information, **sequence correlation factors** θ_λ are calculated as:

$$\theta_\lambda = \frac{1}{L-\lambda} \sum_{k=1}^{L-\lambda} \theta(p_k, p_{k+\lambda}) \quad (4)$$

Where:

- λ is the lag parameter (distance between amino acids),
- $\theta(p_k, p_{k+\lambda})$ is the correlation function between amino acids at positions k and $k + \lambda$, defined as the normalized difference in selected physicochemical properties.

Finally, the PSCAAC feature vector is constructed as:

$$= \left[\frac{AAC_1}{1 + w \sum_{\lambda=1}^{\Lambda} \theta_\lambda}, \dots, \frac{AAC_{20}}{1 + w \sum_{\lambda=1}^{\Lambda} \theta_\lambda}, \frac{w\theta_1}{1 + w \sum_{\lambda=1}^{\Lambda} \theta_\lambda}, \dots, \frac{w\theta_\Lambda}{1 + w \sum_{\lambda=1}^{\Lambda} \theta_\lambda} \right] \quad \text{PSCAAC} \quad (5)$$

With BFDPC and PSCAAC every AFP sequence is mapped to a high-dimensional feature vector that includes compositional and sequence-order information, as well as physicochemical properties. Machine learning classification of antifreeze proteins is subsequently done using these features.

Evaluation Metrics

The performance of the proposed model was evaluated by using some standard measures, such as accuracy, sensitivity (recall), specificity, AUC, and Matthew Correlation Coefficient (MCC). These actions allow forming a good base of evaluating the model handling the classification tasks, in particular in the case when the data are uneven.

Accuracy is actually the rate of samples that are categorized correctly by the model of both AFPs and non-AFPs. It is defined in Eq. (6):

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (6)$$

Here, TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives. Although accuracy gives a general picture of performance, it is not always reliable when class distributions are uneven.

Sensitivity, or recall, measures how effectively the model identifies positive samples (AFPs). It helps reduce the risk of missing true positives and is particularly important in biological applications where undetected DBPs may have serious implications. Sensitivity is calculated in Eq. (7):

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (7)$$

Specificity is used to determine how the model is able to identify negative samples (non- AFPs) correctly. It aims at minimizing false positives and is provided in the form of the equation (8):

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (8)$$

High specificity can be useful in those cases when false positive can result in the unnecessary downstream analysis or interpretation, like filtering proteins during drug development.

The evaluation of MCC is balanced since it takes into consideration all the four outcomes (TP, TN, FP, and FN). It is also extensively applied to unbalanced datasets as it generalises overall performance to a single value. MCC is defined in Eq. (9):

$$MCC = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TN + FN)(TN + FP)(TP + FP)(TP + FN)}} \quad (9)$$

MCC values range from -1 to $+1$, where $+1$ indicates perfect prediction, 0 indicates random performance, and -1 indicates complete disagreement between predictions and true labels.

Taken together, these evaluation metrics give a detailed picture of the model's classification behavior. By examining sensitivity, specificity, AUC, and MCC alongside accuracy, the study provides a well-rounded assessment of the model's stability and effectiveness across different datasets and testing conditions.

Experimental Results and Discussion Performance of the Proposed LightGBM Method

We proposed a prediction model, based on LightGBM, which should enhance the accuracy and reliability of the classification, in the present study. In order to show the effectiveness of the proposed method, we compared LightGBM to numerous popular classifiers, such as XGBoost and AdaBoost. Each of the models was tested in the same experimental conditions to have a reasonable and equal comparison. There was a balanced dataset consisting of positive and negative samples in equal ratios. This dataset was divided into two, 80 percent to be used in training the model with five-fold cross-validation and the remaining 20 percent to be used as an independent test set. The performance of each model was evaluated using the independent data, after training, and compared using major measures like accuracy, sensitivity, specificity, MCC, and AUC. As Table 1 indicates, the suggested LightGBM approach is always competitive and has good generalization abilities on the independent test set. The advantages of LightGBM are that it can capture the nonlinear, complex patterns by the use of gradient

boosting and is also fast to train and can also deal with interactions among features. The proposed LightGBM method is more stable and robust than traditional machine learning models as assessed through evaluation. These results prove that LightGBM is a stable and effective option in this assignment and provides a firm compromise between the predictive power and performance.

Test of Classifier Performance on AFP Prediction

Three machine learning classifiers, including Adaboost, XGBoost, and LightGBM were used to evaluate the performance of two feature sets, BFDPC-AFPs, and PSCAAC-AFPs. The measurement scores were sensitivity (Sen), specificity (Spec), Matthew correlation coefficient (MCC), accuracy (ACC), and area under ROC curve (AUC) that involve an overall assessment of predictive strength of each model. Table 2 summarizes the results.

BFDPC-AFPs Feature Set

In the case of the BFDPC-AFPs feature set, LightGBM had the best predictive performance, with a sensitivity of 97.34, specificity of 97.02, MCC of 0.9468, an accuracy of 97.18 and an AUC of 0.9914. This implies that, LightGBM is able to accurately recognize AFP sequences and reduce the number of false positives. XGBoost was also highly performing with an AUC of 0.9821 and accuracy of 95.14 but Adaboost had a lower specificity (88.17) but a high sensitivity (94.52) and AUC (0.9578). The findings indicate that BFDPC-AFPs can be used to obtain the discriminative features needed to predict AFP.

PSCAAC-AFPs Feature Set

LightGBM was once again better with PSCAAC-AFPs with the best performance characteristics of sensitivity of 97.41, specificity of 97.08, MCC of 0.9441, accuracy of 97.26 and AUC of 0.9892. XGBoost demonstrated similar performance with an accuracy of 95.47% and AUC of 0.9867 when adaboost gave a moderate performance with a sensitivity of 93.47, a specificity of 91.62 and an AUC of 0.9659. The results provide support to the validity of PSCAAC-AFPs in the differentiation of AFPs and non-AFPs.

LightGBM classifier showed consistency and was stronger in the prediction of antifreeze proteins (AFPs) in all the features sets. The discrimination power in the high AUC values is effective whereas the sensitivity and specificity relationship of the models are high implying that the models have successfully balanced false positives and false negatives. Both methods of extraction of features were successfully used: BFDPC-AFPs slightly preferred LightGBM, and PSCAAC-AFPs also demonstrated a high predictive ability. Such findings suggest that the use of ensemble classifiers in conjunction with sophisticated feature descriptors is a stable way of identifying accurate AFP.

Table 2. Performance metrics of different classifiers using BFDPC-AFPs and PSCAAC-AFPs feature sets

Feature Type	Classifier	Sen (%)	Spec (%)	MCC	ACC (%)	AUC
BFDC-AFPs	Adaboost	94.52	88.17	0.8231	91.34	95.78
	XGBoost	96.11	94.08	0.9035	95.14	96.21
	LightGBM	97.34	97.02	0.9468	97.18	98.14
PSCAAC-AFPs	Adaboost	93.47	91.62	0.8534	92.54	96.59
	XGBoost	95.66	95.27	0.9139	95.47	97.67
	LightGBM	97.41	97.08	0.9441	97.26	98.92

Performance Comparison of Classifiers Using BFDC and PSCAAC Feature Descriptors

In figure 2, there is a comparison of the performance between the classifier using two feature descriptors that are BFDC and PSCAAC. There were four metrics of evaluation: accuracy, sensitivity, specificity, and MCC. There are three subplots that compare the three classifiers that were trained using the BFDC features to the classifiers that were trained using the PSCAAC features.

LightGBM has the best performance in both sets of features with a marginal difference in favor of PSCAAC across the accuracy and sensitivity. XGBoost also works well in both scenarios and AdaBoost demonstrates relatively low values, yet it is also competitive.

Similar is the case with the specificity results. LightGBM once again records the largest values in both types of features and PSCAAC features only give a minor improvement over BFDC. XGBoost can be used with high performance, and AdaBoost is rather specific.

These trends are supported by the MCC comparison. The highest MCC of both sets of features is presented by LightGBM which means that the model has a good balance between positive and negative predictions. XGBoost is also trailing closely, and the results of the predictions are enhanced by PSCAAC features. AdaBoost experiences minimum MCC yet it does not ignore PSCAAC.

Generally, these findings indicate that the two feature extraction techniques are effective, with PSCAAC offering a steady performance improvement to all the classifiers. LightGBM provides the most stable and correct predictions of all the models compared.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

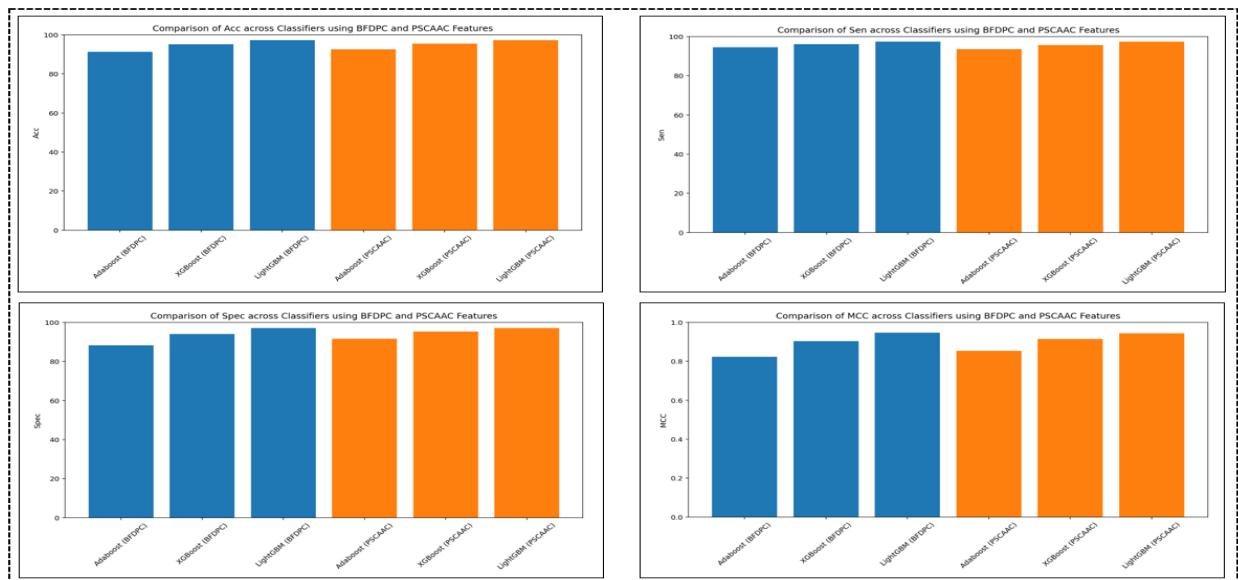


Figure 2. Classification Performance Across BFDC and PSCAAC Features Using AdaBoost, XGBoost, and LightGBM

Comparison of ROC Curves for Machine Learning Models

Figure 3 provides a comparison of the ROC curve of three machine-learning models AdaBoost, XGBoost, and LightGBM across two independent evaluation settings. In the two panels, the ROC curve shows the association between the actual positive rate and the false positive rate in a series of decision levels. A highly-performing model remains nearer to the upper-left corner of the plot, which means that there is a high rate of detection and a small number of false alarms. Compared to AdaBoost, XGBoost and LightGBM are better in terms of their overall separation performance in the left panel. LightGBM has got the maximum AUC value (0.98), closely succeeded by XGBoost (0.97). On the right hand side, the same trends are witnessed. Once again, LightGBM has the highest AUC (0.98), with AUC values of 0.96 and 0.95 by XGBoost and AdaBoost respectively. All the ROC curves are constantly above the diagonal reference line hence all the models are superior to a random classification. In general, LightGBM demonstrates high classification practices in both test environments, succeeded by XGBoost. AdaBoost gives slightly lower results yet gives stable and predictive results.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

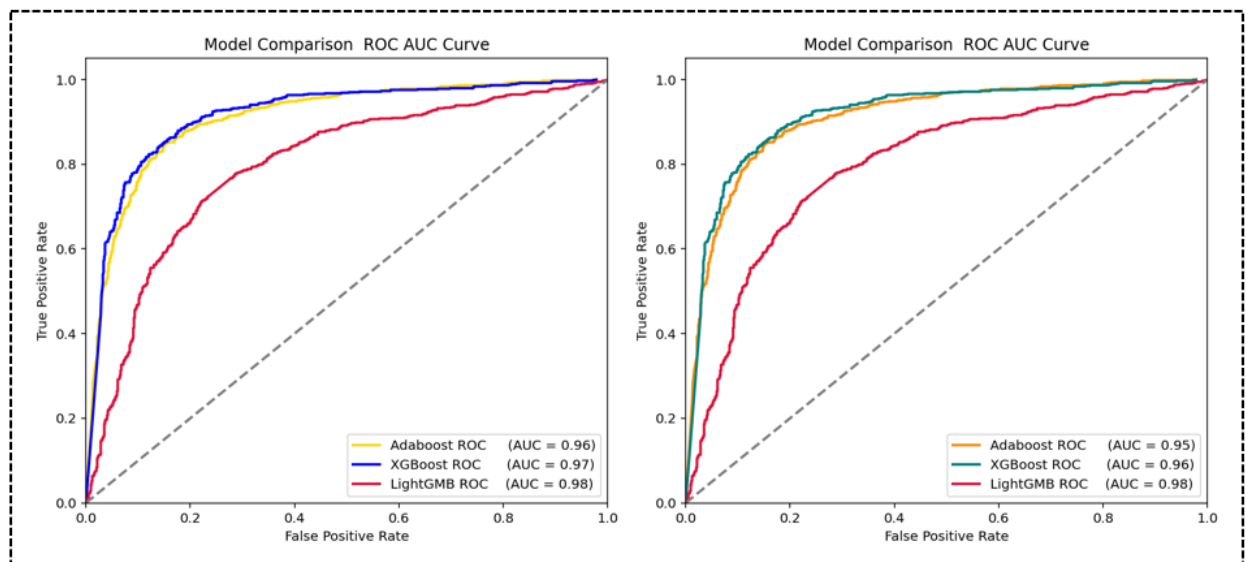


Figure 3. ROC curve comparison of AdaBoost, XGBoost, and LightGBM models across two evaluation settings.

Conclusions

Antifreeze proteins nowadays have become a promising research area due to their broad spectrum of biological applications and their possible applications in biotechnology, food preservation, and cryopreservation, as well as in environmental adaptation research. This amplifies the need to have better strategies that have the ability to capture better sequence patterns and functional signals.

This paper has discussed the current state of development in AFP prediction and has reviewed the role of feature engineering, machine learning, and deep learning in improvements in model performance.

In the future, advances in AFP identification will rely on the creation of more expressive feature descriptors, the creation of larger and better-quality benchmark datasets, and the incorporation of evolutionary, structural, and contextual information. Most of the existing limitations might be resolved through the use of large pretrained language models, multitask learning, and data-efficient training. In order to fairly compare the computational models and to have knowledge of the biological relevance of the predicted AFPs, more systematic experimental validation and standardized evaluation protocols will be required.

Future Directions

Antifreeze proteins prediction is still a divisive field of research in bioinformatics due to their significant roles in various biological processes. Despite the fact that the existing computational processes make valuable forecasts, it remains necessary to have models capable of making more precise distinctions between AFPs. Greater adoption of advanced technologies can be utilized in the future. Modern representation models like BERT, fast Text and biological sub-word models can be used to enhance feature

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

extraction. Such methods can obtain more detailed sequence patterns and can possibly result in more accurate AFP characterization. The combination of deep learning architectures, use of larger training data, and the addition of structural or evolutionary data could also further enhance the results.

Reference

- Ali, F., Akbar, S., Ghulam, A., Maher, Z. A., Unar, A., & Talpur, D. B. (2021). AFP-CMBPred: Computational identification of antifreeze proteins by extending consensus sequences into multi-blocks evolutionary information. *Computers in Biology and Medicine*, 139, 105006.
- Gilbert, J. A., Davies, P. L., & Laybourn-Parry, J. (2005). A hyperactive, Ca²⁺-dependent antifreeze protein in an Antarctic bacterium. *FEMS Microbiology Letters*, 245(1), 67-72.
- Lopes, J. C., Kinasz, C. T., Luiz, A. M. C., Kreusch, M. G., & Duarte, R. T. D. (2024). Frost fighters: unveiling the potential of microbial antifreeze proteins in biotech innovation. *Journal of Applied Microbiology*, 135(6), 1xae140.
- Tas, R. P., Hendrix, M. M., & Voets, I. K. (2022). Direct evidence for pinning of single, ice-bound antifreeze proteins by subzero nanoscopy. *bioRxiv*, 2022-04.
- Mehra, R., Kumar, H., Rafiq, S., Kumar, N., Buttar, H. S., Leicht, K., ... & Korzeniowska, M. (2022). Enhancing yogurt products' ingredients: preservation strategies, processing conditions, analytical detection methods, and therapeutic delivery an overview. *PeerJ*, 10, e14177.
- Libbrecht, K. G. (2017). Physical dynamics of ice crystal growth. *Annual Review of Materials Research*, 47(1), 271-295.
- Deng, G. (1997). Isolation and characterization of a new type of antifreeze protein, from the longhorn sculpin myoxocephalus octodecimspinosus. Boston University.
- Ding, X., Zhang, H., Chen, H., Wang, L., Qian, H., & Qi, X. (2015). Extraction, purification and identification of antifreeze proteins from cold acclimated malting barley (*Hordeum vulgare* L.). *Food chemistry*, 175, 74-81.
- Xu, X., Liu, X., Yu, S., Wang, T., Li, R., Zhang, Y., & Liu, Y. (2024). Medicinal and edible polysaccharides from ancient plants: extraction, isolation, purification, structure, biological activity and market trends of sea buckthorn polysaccharides. *Food & Function*, 15(9), 4703-4723.
- Rives, N., Lamba, V., Cheng, C. C., & Zhuang, X. (2024). Diverse origins of near-identical antifreeze proteins in unrelated fish lineages provide insights into evolutionary mechanisms of new gene birth and protein sequence convergence. *Molecular biology and evolution*, 41(9), msae182.
- Wanigasekara, C., Oromiehie, E., Swain, A., Prusty, B. G., & Nguang, S. K. (2021). Machine learning-based inverse predictive model for AFP based thermoplastic composites. *Journal of Industrial Information Integration*, 22, 100197.
- Chen, S., Zheng, P., Zheng, L., Yao, Q., Meng, Z., Lin, L., ... & Liu, R. (2025). BERT-DomainAFP: Antifreeze protein recognition and classification model based on BERT and structural domain annotation. *iScience*, 28(4).

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

- Fuller, B. J. (2004). Cryoprotectants: the essential antifreezes to protect life in the frozen state. *CryoLetters*, 25(6), 375-388.
- Liu, X., Peng, H., Xie, J., Hu, Y., Liu, F., Wang, X., ... & Tan, S. (2021). Methods in biosynthesis and characterization of the antifreeze protein (AFP) for potential blood cryopreservation. *Journal of Nanomaterials*, 2021(1), 9932538.
- Naing, A. H., & Kim, C. K. (2019). A brief review of applications of antifreeze proteins in cryopreservation and metabolic genetic engineering. *3 Biotech*, 9(9), 329.
- Chi Fai Cheung, R., Bun Ng, T., & Ho Wong, J. (2017). Antifreeze proteins from diverse organisms and their applications: an overview. *Current Protein and Peptide Science*, 18(3), 262-283.
- Rajeev Kumar, S., Kiruba, R., Balamurugan, S., Cardoso, H. G., Birgit, A. S., Zakwan, A., & Sathishkumar, R. (2014). Carrot antifreeze protein enhances chilling tolerance in transgenic tomato. *Acta Physiologiae Plantarum*, 36(1), 21-27.
- Payne, S. R., & Young, O. A. (1995). Effects of pre-slaughter administration of antifreeze proteins on frozen meat quality. *Meat Science*, 41(2), 147-155.
- Jung, W., Kim, Y. P., & Jin, E. (2020). Antifreeze protein-covered surfaces. In *Antifreeze Proteins Volume 2: Biochemistry, Molecular Biology and Applications* (pp. 307-326). Cham: Springer International Publishing.
- Cheng, C. H. C., & Zhuang, X. (2020). Molecular origins and mechanisms of fish antifreeze evolution. *Antifreeze proteins volume 1: environment, systematics and evolution*, 275-313.
- Dhandapani, K., Sivarajan, K., Ravindhiran, R., & Sekar, J. N. (2022). Fungal infections as an uprising threat to human health: Chemosensitization of fungal pathogens with afp from *aspergillus giganteus*. *Frontiers in Cellular and Infection Microbiology*, 12, 887971.
- Ju, Y., Yang, Y., Tang, Q., Wang, M., Zeng, Y., Zhang, Z., ... & Li, L. (2023). Fluorometric detection of alpha-fetoprotein based on the use of a novel organic compound with AIE activity and aptamer-modified magnetic microparticles. *Analytica Chimica Acta*, 1278, 341692.
- Drori, R., Davies, P. L., & Braslavsky, I. (2015). Experimental correlation between thermal hysteresis activity and the distance between antifreeze proteins on an ice surface. *RSC Advances*, 5(11), 7848-7853.
- Estudillo, E., Jiménez, A., Bustamante-Nieves, P. E., Palacios-Reyes, C., Velasco, I., & López-Ornelas, A. (2021). Cryopreservation of gametes and embryos and their molecular changes. *International journal of molecular sciences*, 22(19), 10864.
- Robles, V., G. Valcarce, D., & F. Riesco, M. (2019). The use of antifreeze proteins in the cryopreservation of gametes and embryos. *Biomolecules*, 9(5), 181.
- Khan, A., Uddin, J., Ali, F., Banjar, A., & Daud, A. (2023). Comparative analysis of the existing methods for prediction of antifreeze proteins. *Chemometrics and Intelligent Laboratory Systems*, 232, 104729.
- Jiang, Y., Tong, G., Yin, H., & Xiong, N. (2019). A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters. *IEEE Access*, 7, 118310-118321.

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Feng, C., Wei, H., Li, X., Feng, B., Xu, C., Zhu, X., & Liu, R. (2024). A stacking-based algorithm for antifreeze protein identification using combined physicochemical, pseudo amino acid composition, and reduction property features. *Computers in Biology and Medicine*, 176, 108534.

Zhang, D., & Gong, Y. (2020). The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure. *Ieee Access*, 8, 220990-221003.