Current Research in Food Science 9 (2024) 100842

Contents lists available at ScienceDirect

Current Research in Food Science

journal homepage: www.sciencedirect.com/journal/current-research-in-food-science

Machine learning tools for peptide bioactivity evaluation – Implications for cell culture media optimization and the broader cultivated meat industry

Kathy Sharon Isaac [a], Michelle Combe [a], Greg Potter [b], Stanislav Sokolenko [a,*]

[a] Process Engineering and Applied Science, Dalhousie University, 5273 DaCosta Row, PO Box 15000, Halifax, B3H 4R2, NS, Canada
[b] Lipiferm Scientific, Halifax, NS, Canada

ARTICLE INFO

Handling Editor: Professor A.G. Marangoni

Original content: Machine Learning Tools for Peptide Bioactivity Evaluation Implications for Cell Culture Media Optimization and the Broader Cultivated Meat Industry (Original data)

Keywords:
Peptide bioactivity
Machine learning
Bioactivity classifiers
Hydrolysate peptides
Cell culture media optimization

ABSTRACT

Although bioactive peptides have traditionally been studied for their health-promoting qualities in the context of nutrition and medicine, the past twenty years have seen a steady increase in their application to cell culture media optimization. Complex natural sources of bioactive peptides, such as hydrolysates, offer a sustainable and cost-effective means of promoting cellular growth, making them an essential component of scaling-up cultivated meat production. However, the sheer diversity of hydrolysates makes product selection difficult, highlighting the need for functional characterization. Traditional wet-lab techniques for isolating and estimating peptide bioactivity cannot keep pace with peptide identification using high-throughput tools such as mass spectrometry, requiring the development and use of machine learning-based classifiers.

This review provides a comprehensive list of available software tools to evaluate peptide bioactivity, classified and compared based on the algorithm, training set, functionality, and limitations of the underlying models. We curated independent test sets to compare the predictive performance of different models based on specific bioactivity classification relevant to promoting cell culture growth: antioxidant and anti-inflammatory. A comprehensive screening of all bioactivity classifiers revealed that while there are approximately fifty tools to elucidate antimicrobial activity and sixteen that predict anti-inflammatory activity, fewer tools are available for other functionalities related to cell growth — five that predict antioxidant activity and two for growth factor and/or cell signaling prediction. A thorough evaluation of the available tools revealed significant issues with sensitivity, specificity, and overall accuracy. Despite the overall interest in estimating peptide bioactivity, our work highlights key gaps in the broader adoption of existing software for the specific application of cell culture media optimization in the context of cultivated meat and beyond.

1. Introduction

The production of cultivated meat via cell culture is a developing technology with the potential to sustainably address increasing worldwide demand for meat. In contrast to traditional animal sources, cultivated meat is projected to emit fewer greenhouse gases and utilize fewer resources, including energy, land, and water (Tuomisto and Mattos, 2011; Carus et al., 2019). Furthermore, animal-free meat avoids the ethical concerns surrounding animal rearing, and can address more recent health concerns over the transmission of animal diseases and zoonotic epidemics (such as COVID-19) (O'Neill et al., 2021). Despite these advantages, one of the biggest constraints in the economic feasibility and productivity of cultivated meat lies in the cost and composition of cell culture media (O'Neill et al., 2021). Cellular proliferation in vitro is sustained by a complex cocktail of basic macronutrients as well as other essential components, such as growth factors, vitamins, and hormones (Chandra et al., 2022; Siemensma et al., 2010). In traditional mammalian cell culture, these requirements are often met through the addition of fetal bovine serum (FBS) (Siemensma et al., 2010; Obaidi et al., 2021), which is not a viable option for true animal-free status. More broadly, FBS usage as a cell culture media supplement is regarded as unsustainable due to the risk of microbial contamination, large

Abbreviations: ML, Machine learning; DL, deep learning; RF, random forest; SVM, support vector machine; CNN, convolutional neural network; AMP, antimicrobial peptide; ABP, antibacterial peptide; AFP, antifungal peptide; APP, antiparasitic peptide; ROC, receiver operating curve; AUC, area under the receiver operating curve; FBS, fetal bovine serum.

* Corresponding author.
E-mail address: Stanislav.Sokolenko@dal.ca (S. Sokolenko).

https://doi.org/10.1016/j.crfs.2024.100842
Received 31 May 2024; Accepted 7 September 2024
Available online 16 September 2024

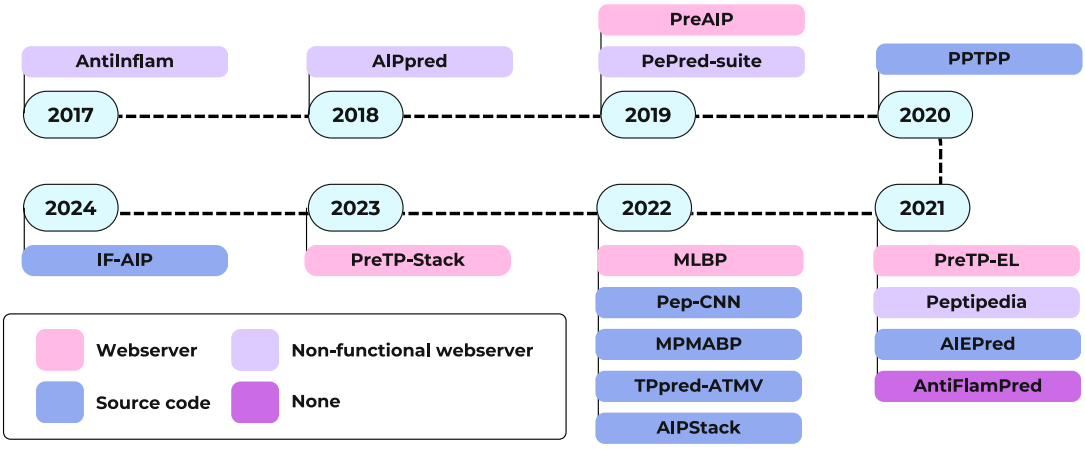

Fig. 1. Chronology of anti-inflammatory classifiers.

production costs, and high batch-to-batch variability, all of which compromise product quality and consistency (Ho et al., 2021; Obaidi et al., 2021). The development and adoption of low-cost, food grade cell culture media is thus necessary to produce affordable cultivated meat (Humbird, 2021).

Plant protein hydrolysates have emerged as a cost-effective, ethical, and potentially more sustainable alternative to conventionally produced amino acid components and animal serum (Humbird, 2021; Ho et al., 2021). Hydrolysates were initially developed from meat and microorganism sources such as yeast, but have since expanded to include a variety of different plant substrates such as cotton, soy, wheat, and rice (Ho et al., 2021; Obaidi et al., 2021). The diversity in hydrolysate products makes their compositional and functional characterization key to efficient media optimization, by mapping the hydrolysate composition to the unique metabolic requirements of the cell type of interest (Humbird, 2021). Hydrolysates are primarily peptide-rich products (Ho et al., 2021; Obaidi et al., 2021; Djemal et al., 2021), many of which are considered "bioactive", i.e., believed to play a complex role in cellular metabolism (Siemensma et al., 2010). Bioactive peptides are peptide sequences that exert physiological effects as growth factors, antioxidants, signalling molecules, and metal transporters (Ho et al., 2021; Kumar et al., 2022). Plant hydrolysate peptide bioactivity has traditionally been studied in the context of human health and nutrition, as they have been linked to possessing anti-diabetic, anti-cancer, anti-microbial and anti-hypertensive properties, however, the last twenty years of research has generated increasing appreciation for the role of peptide bioactivity in the context of culture media (Dayem et al., 2023; Spearman et al., 2014).

The identification of novel bioactive peptides remains challenging as it depends on peptide sequence, structure, and a host of physicochemical properties (Siemensma et al., 2010; Kumar et al., 2022). Although experimental validation is the only gold standard for estimating bioactivity (Farias et al., 2023; de Castro and Sato, 2014; Zhang et al., 2023; Hsieh et al., 2022), it remains prohibitively resource intensive to validate thousands of peptides per hydrolysate product. An alternative approach that has emerged over the past 15 years is *in silico* bioactivity classification, performed primarily using supervised learning models trained on known bioactive peptides. These models use machine learning (ML) algorithms to identify complex patterns in the primary sequence, physiochemical properties and/or structural properties of the labelled training set to predict the bioactivity of unknown peptides (Bárcenas et al., 2022).

Bioactivity prediction tools emerged as early as 2010 with the introduction of the antibacterial peptide (ABP) classifier AntiBP2 (Lata et al., 2010). This was soon followed by several antimicrobial peptide (AMP) classifiers, Wang et al. (2011), ANFIS (Fernandes et al., 2012) and CS-AMPpred (Porto et al., 2012), as well as a general bioactivity classifier PeptideRanker (Mooney et al., 2012), which were published between 2011 and 2012. Since then, both the number of classifiers and bioactivity categories have steadily expanded. Newer models aim to improve accuracy by using more advanced algorithms, larger and/or more accurate training sets and superior features. The growing interest in therapeutic peptides has also led to the rise of classifiers for anti-inflammatory (Gupta et al., 2017), cell-penetrating (CPP) (Holton et al., 2013; Pandey et al., 2018; de Oliveira et al., 2021), anti-hypertensive (Manavalan et al., 2019), and anti-cancer (Saravanan and Lakshmi, 2015; Schaduangrat et al., 2019; Ahmed et al., 2021) peptide prediction. This expansion in bioactivity prediction models complicates model selection, necessitating comparative studies. Xu et al. (2021) performed an excellent comparison of 28 AMP classifiers using independent test sets and validation sets but the analysis did not extend to other categories of classifiers. Independent test sets were also used by Su et al. (2020) and Manavalan et al. (2022) to objectively evaluate the predictive performance of CPP and anti-SARS-CoV-2 peptides, respectively. However, to the best of our knowledge, no such independent comparison has been performed for other important categories more relevant to cell culture media optimization, including antioxidant and anti-inflammatory peptide classifiers.

The objectives of this review are twofold: first, to characterize all classifiers available for the prediction of bioactivity categories relevant to media optimization (including antioxidant, anti-inflammatory, antimicrobial, and growth factor); and second, to evaluate classifier performance for bioactivity categories with three or more classifiers using independent test sets. General characterization is aimed at comparing classifier properties such as general architecture, e.g., algorithms employed, features selected, and the size and properties of the training sets, as well as their functionality, e.g., webserver availability, to assist classifier selection. Meanwhile, independent evaluation of classifier

Table 1
Summary of anti-inflammatory classifiers.

Name	Reference	Algorithm	Positive database	Positive peptides	Negative peptides	Data availability	Offline code	Webserver	Citations (all, articles)
AntiInflam	Gupta et al. (2017)	SVM	IEDB	1258	1887	No	No	Inaccessible	89, 66
AIPpred	Manavalan et al. (2018)	RF	IEDB	1258	1887	No	No	Inaccessible	147, 114
PreAIP	Khatun et al. (2019)	RF	AIPpred and IEDB	1258	1887	Yes	No	Yes	88, 68
PEPred-suite	Wei et al. (2019)	RF	AIPpred	1258	1887	No	No	Inaccessible	117, 98
PPTPP	Zhang and Zou (2020)	RF	AIPpred	1258	1887	No	Yes	No	68, 59
AntiFlamPred	Alotaibi et al. (2021)	CNN	AIPpred, PreAIP, PEPred-suite, and IEDB	1911	4240	No	No	No	4, 4
AIEPred	Zhang et al. (2021a)	RF	IEDB and Proinflam	690	1009	Yes	Yes	No	30, 26
PreTP-EL	Guo et al. (2021)	Ensemble (SVM and RF)	AIPpred	1258	1887	Yes	No	Yes	28, 23
Peptipedia	Quiroz et al. (2021)	RF	UniProt, LAMP2, SATPdb, DBAASP, DRAMP, CAMP, etc.	–	–	No	No	Inaccessible	14, 8
AIPStack	Deng et al. (2022)	Ensemble (ET and RF)	IEDB	1493	2276	Yes	Yes	No	8, 6
Pep-CNN	Zhang and Li (2022)	CNN	AIPpred	1258	1887	Yes	Yes	No	8, 5
MPMABP	Li et al. (2022b)	CNN and Bi-LSTM	MLBP	1342	–	Yes	Yes	No	7, 6
TPpred-ATMV	Yan et al. (2022)	AMVTLF	AIPpred	–	–	No	Yes	No	29, 26
MLBP	Tang et al. (2022)	CNN	PreAIP	1342	–	Yes	Yes	Yes	31, 25
PreTP-Stack	Yan et al. (2023)	Stacked ensemble (SVM, RF, LDA, XGB, and AMV)	AIPpred	1258	1887	Yes	No	Yes	13, 10
IF-AIP	Gaffar et al. (2024)	RF, LGBM, XGB, ETC, and CatBoost	AntiInflam and iAIPs	1451	2339	Yes	Yes	No	4, 3

Algorithm names: random forests, RF; support vector machine, SVM; convolutional neural networks, CNN; bidirectional long short term memory, Bi-LSTM; adaptive multi-view based on the tensor learning framework, AMVTLF; light gradient boost machine, LGBM; extreme gradient boositing, XGB; extra tree classifier, ETC; linear discriminant analysis, LDA; auto-weighted multi-view learning, AMV.

performance using common test sets augments general characterization with a better estimate of accuracy (than possible based on self-reported statistics) and can serve to guide future application and development. From this, we aim to identify potential areas of improvement to advance the use of bioactivity classification for the rational selection of hydrolysate products in media optimization.

2. Classifier architecture

2.1. Anti-inflammatory

Anti-inflammatory peptides are seeing increased interest within the context of peptide-based therapies to combat autoimmune and inflammatory illnesses, such as Alzheimer's and rheumatoid arthritis (Gupta et al., 2017). These peptides function to lower the release of prostaglandins, thromboxane, pro-inflammatory cytokines, etc. by cells, which leads to reduced swelling and redness (Gupta et al., 2017). Their use also extends to cell culture studies where they have been shown to lower the release of reactive oxygen species (ROS), nitric oxide and pro-inflammatory cytokines when added to lipopolysaccharide stimulated culture of macrophages (Lee et al., 2015) and hepatic cell lines (Cruz-Chamorro et al., 2022), thereby enhancing cell growth. To help accelerate the search of novel anti-inflammatory peptides, 16 anti-inflammatory classifiers have been published over the past 8 years as presented in Fig. 1, with AntiInflam (Gupta et al., 2017) pioneering this bioactivity class. A summary of these classifiers is shown in Table 1.

While all tools prior to 2021 utilized traditional ML algorithms, predominantly based on the random forest (RF) methodology, an increasing number of models have since incorporated deep learning (DL) algorithms, particularly convolutional neural networks (CNNs). In CNN models, the input is typically a sequence represented as a one-hot encoded matrix.[1] The model uses filters to extract features from this input, which are then used to predict the classification output (Xu et al., 2021). In contrast, RF models are more simplistic, built upon an ensemble of decision trees applied to user-defined features (Sarker, 2021). Although DL based models are more complex than traditional ML models, they are not necessarily superior, as they are more susceptible to overfitting, especially when trained on small datasets, which can compromise the generalizability and accuracy of the model (Bejani and Ghatee, 2021).

The datasets used to train the classifiers are generally small, with most containing only 1000–2000 anti-inflammatory peptides, and slightly imbalanced, having more negative peptides than positive peptides. Such data imbalances are known to introduce bias towards the larger category, potentially resulting in an increased number of false negatives (Lertampaiporn et al., 2021). Several classifiers, AntiInflamPred (Alotaibi et al., 2021), TPpred-ATMV (Yan et al., 2022) and Peptipedia (Quiroz et al., 2021), did not make their datasets available or

[1] One-hot encoding is a numerical representation of a protein or peptide sequence of length m as an m x n matrix, where n represents the total number of all possible amino acids. Each row represents a residue from the protein or peptide with the corresponding column matching the residue containing a 1 and all other columns in that row containing a 0.

Table 2
Summary of user interface of functional anti-inflammatory webservers.

Name	Input type	Maximum input	Results download	Additional options
PreAIP	FASTA sequences	–	No	Can retrieve results using jobID
MLBP	FASTA sequences or FASTA file	1000 sequences	CSV	Email notification
PreTP-EL	FASTA sequences or FASTA file	1000 sequences	ZIP (text file)	Email notification and user determined threshold
PreTP-Stack	FASTA sequences or FASTA file	–	CSV	Email notification and user determined threshold

use existing datasets. Excepting these, 5354 unique peptides were used to train the classifiers.[2] These peptides ranged in length from 5 to 50 amino acid residues, although most of these were on the shorter side of this range, with a mode of 15 and average of 16.5 residues. Many of the training set were found to have considerable overlap due to the fact that most peptides were derived from the IEDB database (Vita et al., 2019), with 85% of the peptides found to be shared by two or more classifiers. As the choice of training set strongly impacts model accurary, this level of overlap suggests that classifier performance is expected to be similar across the classifiers, with most differences stemming from choice of algorithm and features. For instance PreAIP (Khatun et al., 2019), AIPpred (Manavalan et al., 2018), PPTPP (Zhang and Zou, 2020), PEPred-Suite (Wei et al., 2019), PreTP-Stack (Yan et al., 2023), PreTP-EL (Guo et al., 2021), Pep-CNN (Zhang and Li, 2022), and Anti-Inflam were trained using the same dataset and all but Pep-CNN were trained using different ML algorithms, all of which resulted in similar self reported model performance scores. In contrast, AIPStack (Deng et al., 2022) stands out as it uses one of largest and the most distinct training set containing 684 unique peptides.

Approximately half of the published studies included a public webserver, with the rest providing offline source code. Although webservers appear more popular than offline code (as estimated from the citation count of the associated journal article), a number of these are no longer accessible (including all three published before 2020). Indeed, AIPpred and PEPred-suite are among the most highly cited tools despite the fact that they are currently unavailable. A summary of the user interface of the functional webservers is seen in Table 2. Of these, PreTP-Stack stands out from the rest based on its ability to handle longer peptide sequences, process more than 1000 sequences per request, accept both text or file inputs, and download results as a single CSV file.

2.2. Antioxidant

Antioxidant peptides are peptides that neutralize free radicals and reactive oxygen species (ROS). They have been studied for their therapeutic potential in combating ailments that increase oxidative stress such as diabetes and hypertension (Sun et al., 2021). In cell culture, increased oxidative stress can lead to cell death and lower cell concentrations, with antioxidants typically added to culture media to combat stress (Halliwell, 2014). The stability and solubility of antioxidant peptides offer important advantages over traditional additives such as vitamin E and vitamin C (Halliwell, 2014). Glutathione is just one example of a tripeptide that is frequently added to culture media in this capacity (Kwon et al., 2019). All five antioxidant classifiers identified as part of this review have been published within the last four years: AnOxPePred (Olsen et al., 2020), MultiPep (Grønning et al., 2021), Peptipedia, AnOxPP (Qin et al., 2023) and Deep2Pep (Chen et al., 2024). Of these, AnOxPePred and AnOxPP are dedicated classifiers that only predict antioxidant activity, while the others are general classifiers that predict antioxidant activity as just one form of bioactivity. AnOxPePred further breaks down antioxidant activity into two specific categories — free radical scavenger and ion chelator. A summary of these classifiers is shown in Table 3.

All but Peptipedia were constructed using DL algorithms for classification. The majority of peptides utilized for training these classifiers were short, with an average length of 6 residues. AnOxPP and AnOxPePred impose stringent length restrictions of 19 and 30 residues, respectively, to align with the maximum length of peptides in their respective training sets. In contrast, MultiPep has the capability to classify peptides with lengths of up to 200 residues, despite being trained on peptides smaller than 60 residues. Of the 1235 total antioxidant peptides used to train the classifiers, 431 were common to all. AnOxPP used the largest dataset, comprising of 1060 antioxidant peptides to train the classifier while the others used fewer than 700. No overlap was observed with the negative datasets, but it should be noted that we observed 244 contradictory peptides labelled as negative by AnOxPePred but positive by AnOxPP and/or MultiPep. This is likely because AnOxPePred sourced antioxidant peptides solely from BIOPEP-UWM (Minkiewicz et al., 2019) while AnOxPP and MultiPep used additional newer databases such as DFBP (Qin et al., 2022) and DBAASP (Pirtskhalava et al., 2021), respectively. As these databases are independently curated from peptides published in literature at different time periods, and newer databases additionally contain antioxidant peptides that were recently discovered. Moreover, AnOxPePred built a negative dataset using a random peptide generation approach, where proteins that are not known to function as antioxidants are obtained from UniProt and peptides are randomly generated from these protein sequences. As parent protein function is not a reliable indicator of peptide function, this may lead to these “non-antioxidant” peptides actually exhibiting antioxidant activity. For example, the peptide AGTTCLFTPLALPYDYSH, which was randomly generated by AnOxPePred, is not listed as an antioxidant from the BIOPEP-UWM database and was consequently labelled as negative by AnOxPePred, but is listed as an antioxidant in the DFBP database and was thus labelled as positive by AnOxPP. This observation reinforces the need for accurate negative dataset selection as a key consideration in classifier development.

AnOxPePred, AnOxPP and MultiPep are all currently available as functional accessible webservers (Table 4), but differ considerably in the number of input peptides classified per request — AnOxPePred allows 20 peptides, AnOxPP 200 peptides, and MultiPep 500 peptides. While none of the tools provide an API for processing large requests, MultiPep is additionally available as a command-line Python tool. Despite its limitations, the most popular tool appears to be AnOxPePred, with 75 citations to date. Its popularity can likely be attributed to its status as the pioneering antioxidant classifier as well as a number of distinctive features. Not only does it predict two antioxidant categories, but it also possesses the capability to predict antioxidant peptides from parent proteins both exhaustively and with common digestion enzymes such as trypsin.

2.3. Antimicrobial

The most popular category of peptide bioactivity classification is AMP prediction, with 50 tools published over the course of the past 15 years. This popularity is due in part to the potential of AMPs as drugs, since they are well tolerated by humans and are less susceptible to antimicrobial resistance (Xu et al., 2021). AMPs can also find use in cell culture media formulations as the nutrient rich culture medium is a fertile breeding ground for microorganisms which leads to

[2] Analyzing the training set peptides revealed a discrepancy: 521 peptides had contradictory labels as they were labelled as positive anti-inflammatory peptides by some classifiers and negative by others. This discrepancy can be partially attributed to databases growing over time as new peptides are discovered and the tendency for classifiers to use peptides with unknown function as their negative set.

Table 3
Summary of antioxidant classifiers.

Name	Reference	Algorithm	Positive database	Positive peptides	Negative peptides	Data availability	Offline code	Webserver	Citations (all, articles)
AnOxPePred	Olsen et al. (2020)	CNN	BIOPEP-UWM	696	218	Yes	Yes	Yes	75, 59
Peptipedia	Quiroz et al. (2021)	RF	DFBP and BIOPEP-UWM	–	–	No	No	Inaccessible	14, 8
MultiPep	Grønning et al. (2021)	CNN	BIOPEP-UWM, APD3, LAMP2, DBAASP, etc	540	–	Yes	Yes	Yes	15, 11
AnOxPP	Qin et al. (2023)	BiLSTM	UniProt, APD, AHTPDB, DFBP, BIOPEP-UWM and BGI-marine	848	848	Yes	No	Yes	7, 3
Deep2Pep	Chen et al. (2024)	BiLSTM	UniProt, LAMP2, SATPdb, DBAASP, DRAMP, CAMP, etc.	–	–	No	No	No	0, 0

Algorithm names: random forest, RF, convolutional neural networks, CNN; bidirectional long short term memory, Bi-LSTM.

Table 4
Summary of user interface of functional antioxidant webservers.

Name	Sequence length	Input type	Maximum input	Results download	Additional options
AnOxPePred	2–30	FASTA sequences or FASTA file	50 sequences	CSV	Modes: peptide, protein exhaustive and enzymatic digestion
MultiPep	2–200	List of sequences	500 sequences	CSV	Outputs average or maximum of all models
AnOxPP	2–19	FASTA sequences or FASTA file	200 sequences	FASTA file	

contamination of the cell culture (Mahmood and Ali, 2017). AMP is an umbrella term that generally includes antibacterial, antiviral, antifungal, and antiparasitic peptides (Zhang et al., 2021b; Benfield and Henriques, 2020). While some classifiers distinguish between these antimicrobial categories, others aim to predict broad spectrum antimicrobial activity (Bárcenas et al., 2022). AMPs vary in their mode of action, with some acting directly and specifically against the microbe and others stimulating and amplifying the immune response (Benfield and Henriques, 2020). Cationic AMPs, for instance, increase the permeability of and disrupt the negatively charged bacterial cell membrane (Zhang et al., 2021b). Other AMPs specifically interrupt microbial metabolic processes by, for example, inhibiting key protein synthesis, DNA transcription and/or translation, enzyme activity and cell wall synthesis (Zhang et al., 2021b; Benfield and Henriques, 2020). Very few of the existing classifiers focus on classifying a specific class of AMPs, such as cationic AMPs, and none currently aim to predict their mode of action.

2.3.1. Broad spectrum antimicrobial

To date, 34 classifiers have been published that predict broad spectrum AMP bioactivity (Fig. 2), a summary of which is shown in Table 5. While these classifiers claim to detect broad spectrum AMPs, many were trained using datasets with a large number of ABPs over other categories of AMPs. For example, iAMPpred (Meher et al., 2017) was trained using 3417 anti-bacterial peptides (ABPs) but only 739 anti-viral peptides (AVPs). Such data imbalances can lead to these broad-spectrum classifiers being biased towards specific sub-categories, such as ABPs, and thus caution should be exercised while utilizing them for predicting other categories such as AVPs. Two of these classifiers, DefPred (Kaur et al., 2021) and Prediction of Linear Cationic AMPs (Ümmü et al., 2022) are more specific in the type of AMP they predict, with the former predicting defensins (host defense proteins found in plants and animals) and the latter linear cationic AMPs (which work to disrupt the microbial cell wall).

A majority (73%) of the tools are based on traditional ML algorithms, with RF and SVM being most popular, and the rest on more recent DL algorithms, such as CNN. The most cited models include iAMPpred, AmPEP (Bhadra et al., 2018), and Deep-AmPEP30 (Yan et al., 2020), all of which released an accompanying webserver. As noted earlier, the convenience of webserver access is somewhat offset by the need for continuous maintenance, with only 8 out of 19 AMP webservers still accessible as of this writing.

Xu et al. (2021) evaluated the performance of 28 AMP classifiers using an independent test set with 1536 AMPs and 1536 non-AMPs and found that amPEPpy (Lawrence et al., 2021), an RF classifier, performed the best with the highest area under the receiver operating characteristic curve (AUC) score of 0.742. This performance was attributed to the classifier's large and diverse training set of 3268 AMP and non-AMP peptides (Xu et al., 2021; Lawrence et al., 2021). The main drawbacks identified in existing AMP classifiers is a high false positive rate (Xu et al., 2021). This may be due to the fact that while positive peptides are usually obtained from similar databases of peptides with known experimentally validated bioactivity, negative peptides are less consistent, as it is unusual to see experimental validation of non-bioactivity (Bárcenas et al., 2022). These peptides are typically obtained from UniProt with no AMP or related annotation (Bárcenas et al., 2022). However, this approach generally fails to capture the full breadth of non-bioactive peptides and can result in poor performance outside of specific test cases (Sidorczuk et al., 2022).

2.3.2. Specific antimicrobial

Among 22 specific AMP classifiers (Fig. 3), six only predict ABPs, with four of these additionally able to discriminate between gram-positive and gram-negative ABPs. The remaining classifiers predict multiple categories of AMPs, including ABPs, AVPs, antifungal peptides (AFPs), and antiparasitic peptides (APPs). A summary of these classifiers is shown in Table 6. While some earlier tools predicting AVPs, such as ClassAMP (Joseph et al., 2012), used a small training set with only 54 AVPs, newer tools like AMPfun (Chung et al., 2020) utilize much larger datasets, e.g., 1400 AVPs. Of all 22 AMP classifiers, 18 have developed web servers, with 10 still operational. The most highly cited classifiers include ones published first like iAMP-2L (Xiao et al., 2013), a multiple classifier, and more recent ones such as Antimicrobial Peptide Scanner vr.2 (Veltri et al., 2018), which predicts peptides targeting both gram-positive and gram-negative bacteria. The Antimicrobial Peptide Scanner vr.2 webserver is particularly notable for its ability to process up to 50,000 sequences per request, with loose length constraints of 10–200 residues, likely contributing to its increased use.

2.4. Other categories

Beyond the bioactivity classifications discussed thus far, some peptides are believed to promote cell growth by functioning as growth factors, signal peptides and hormones. However, the prediction of these

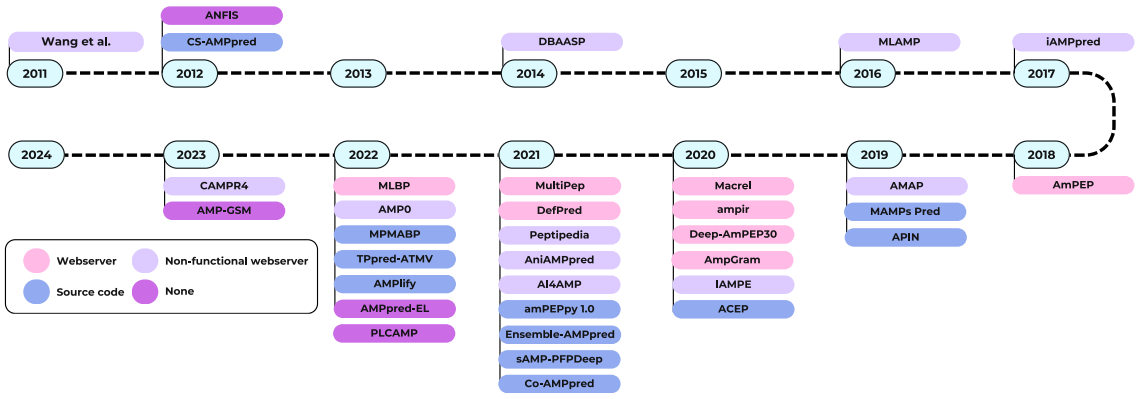

Fig. 2. Chronology of broad spectrum antimicrobial classifiers.

functions is currently limited to only two tools: Peptipedia and MultiPep, of which only MultiPep remains available. The models implemented by MultiPep were trained using a relatively large dataset with 3760 cytokines/growth-factors and 6943 hormones and had a self reported sensitivity of approximately 0.76 for both categories (Grønning et al., 2021).

3. Classifier validation

With the sheer number of tools available for each category of peptide bioactivity and the diversity in training and validation sets, establishing an accurate assessment of predictive performance requires the use of a standardized framework. As antimicrobial classification has already been thoroughly analyzed by Xu et al. (2021), we chose to focus on anti-inflammatory and antioxidant classification for this review.

3.1. Independent test sets

To assess the performance of anti-inflammatory and antioxidant classifiers, we created two independent test sets for each classifier. Independent test set I was constructed by gathering all peptides used to train other classifiers within the same category and then excluding the peptides used to train the classifier being evaluated. Technically, this resulted in a set of test sets, with each classifier assessed using a slightly different variant. Under this strategy, each classifier was essentially tested using the training sets of other classifiers, excluding the peptides that it was originally trained on. The goal of this independent test set was to get a sense of classifier performance by using high confidence peptides. Since anti-inflammatory tools were typically trained on data from the IEDB database and antioxidant tools from the BIOPEP-UWM database, but utilized different sets of peptides, our goal was to evaluate these tools using peptides that were reliable enough to have been part of the training dataset of at least one tool. Furthermore, we excluded 521 anti-inflammatory peptides and 244 antioxidant peptides with contradictory labels from all test sets to ensure the reliability of the evaluations (see sections 2.1 and 2.2). This resulted in a unique independent test set for each classifier, with the peptide breakdown shown in Table 7.

Independent test set II was developed as a truly common external dataset. For anti-inflammatory classifiers, this set of peptides was obtained from Peptipedia, BIOPEP-UWM, and MDPDB (Yang et al., 2023) databases, containing known anti-inflammatory peptides. Negative anti-inflammatory peptides were sourced from UniProt, ensuring that none of the peptides were annotated with GO terms indicating anti-inflammatory properties. Similarly, the peptides for testing the antioxidant classifiers utilized positive peptides from AODB (Deng et al., 2023), an antioxidant peptide database, and Peptipedia, while negative peptides were sourced from UniProt, ensuring absence of antioxidant-related GO annotations such as “antioxidant” and “scavenger”. For both the positive and negative datasets for each class, peptides that were used to train the classifiers were filtered out along with any peptide with non-standard amino acid residues. Subsequently, peptides were filtered by length to comply with the most stringent length constraints accepted by all classifiers. For anti-inflammatory classifiers, this meant a length range of 5–50 residues, while for antioxidant classifiers, it was 2–19 residues. As this resulted in more negative peptides than positive, an equal number of negative peptides were randomly sampled to balance the length distribution between positive and negative sets. This resulted in 350 positive and negative peptides each to evaluate the anti-inflammatory classifiers and 496 positive and negative peptides each to evaluate antioxidant classifiers. All independent test sets are available at https://doi.org/10.5281/zenodo.11402692.

Tested classifiers were limited to those that were available as functional webservers or as functional stand-alone tools. Stand-alone tools were excluded if we were unable to achieve results on a trivial example within 6 h of installation (which may not have been possible due to missing documentation or occasional bugs). Of the 16 anti-inflammatory tools, seven were released as webservers but only four remain functional. Although eight provided their model code as a GitHub repository, all were excluded due to insufficient documentation. Of the five released antioxidant classifiers, only three are available as functional webservers. Of these three, MultiPep is also available as a functional stand-alone command-line interface Python tool, which was used in lieu of the webserver to process large requests. As AnOxPePred outputs both free radical scavenger and ion chelation activities, these were evaluated separately as AnOxPePred-Scavenger and AnOxPePred-Chelator, respectively.

Table 5
Summary of broad spectrum antimicrobial classifiers.

Name	Reference	Algorithm	Positive database	Positive peptides	Negative peptides	Data availability	Offline code	Webserver	Citations (all, articles)
Wang et al. (2011)	Wang et al. (2011)	kNN	CAMP	2752	10014	No	No	Inaccessible	182, 143
ANFIS	Fernandes et al. (2012)	ANFIS	APD2	115 clusters	116 clusters	No	No	No	34, 22
CS-AMPpred	Porto et al. (2012)	SVM	PDB	310	310	No	Yes	No	76, 54
DBAASP	Gogoladze et al. (2014)	Physicochemical	PubMed	1083	–	No	No	Inaccessible	86, 49
MLAMP	Lin and Xu (2016)	SMOTE	iAMP-2L	879	2405	No	No	Inaccessible	77, 56
iAMPpred	Meher et al. (2017)	SVM	CAMP, APD3 and AntiBP2, AVPpred, and LAMP	5652	3261	No	No	Inaccessible	311, 253
AmPEP	Bhadra et al. (2018)	RF	APD3, CAMPR, and LAMP	3268	166791	Yes	No	Yes	165, 116
MAMPs Pred	Lin et al. (2019)	RF	APD	2618	4371	No	No	No	34, 22
AMAP	Gull et al. (2019)	SVM	APD3	2704	5156	No	No	Inaccessible	53, 34
APIN	Su et al. (2019)	DNN	DAMP, AntiBP2, AIPpred, and APD3	6168	13861	Yes	Yes	No	56, 37
AmpGram	Burdukiewicz et al. (2020)	RF	dbAMP	2463	2463	Yes	Yes	Yes	49, 35
Deep-AmPEP30	Yan et al. (2020)	CNN	AmPEP	1529	1529	Yes	Yes	Yes	172, 93
ACEP	Fu et al. (2020)	DNN	AMPScanner		0	Yes	Yes	No	28, 18
ampir	Fingerhut et al. (2020)	SVM	UniProt	–	–	Yes	Yes	Yes	22, 17
IAMPE	Kavousi et al. (2020)	Ensemble	CAMP, LAMP, ADAM, and AntiBP	2667	1390	No	No	Inaccessible	50, 33
Macrel	Santos-Júnior et al. (2020)	RF	ADP3, CAMPR3, and LAMP	1197	1197–60000	Yes	Yes	Yes	30, 22
amPEPpy	Lawrence et al. (2021)	RF	APD3, CAMPR, and LAMP	3268	3268	Yes	Yes	No	45, 28
AI4AMP	Lin et al. (2021)	DNN	APD3, LAMP, CAMP3, and DRAMP	3528	3528	Yes	Yes	Inaccessible	31, 21
DefPred	Kaur et al. (2021)	SVM	DRAMP2.0, and CAMPR3	1036	1036	Yes	Yes	Yes	4, 1
AniAMPpred	Sharma et al. (2021a)	SVM	NCBI and starPepDB	6657	6773	No	No	Inaccessible	34, 22
Co-AMPpred	Singh et al. (2021)	Ensemble	Deep-AmPEP30	1529	1529	Yes	Yes	No	17, 13
Ensemble-AMPPred	Lertampaiporn et al. (2021)	Ensemble	APD, BACTIBASE, BAGEL3, CAMP, DRAMO, DBAASP, BIOPEP-UWM, etc	1800	1800	No	No	No	21, 13
sAMP-PFPDeep	Hussain (2022)	DNN	ADAM	1529	1529	Yes	Yes	No	28, 18
MultiPep	Grønning et al. (2021)	CNN	APD3, BioDADPep, BIOPEP-UWM, CAMPR3, DBAASP, LAMP2, PeptideDB, and SATPdb	14362	–	Yes	Yes	Yes	15, 11
Peptipedia	Quiroz et al. (2021)	RF	UniProt, LAMP2, SATPdb, DBAASP, DRAMP, CAMP, etc.	–	–	No	No	Inaccessible	14, 8
MLBP	Tang et al. (2022)	CNN	PreAIP, mAHTPred, BioDADPep, AntiCP 2.0, and AMPfun	1342	–	Yes	Yes	Yes	31, 25
Prediction of Linear Cationic Antimicrobial Peptides	Ümmü et al. (2022)	RF	DBAASP	396	308	No	No	No	11, 7
AMP0	Gull and Minhas (2022)	SVM	DBAASP vr. 2	5710	–	No	No	Inaccessible	53, 34
AMPlify	Li et al. (2022a)	BiLSTM	APD3 and DADP	3338	3338	Yes	Yes	No	56, 37
AMPpred-EL	Lv et al. (2022)	Ensemble	SATPdb, ADAM, AMPfun, APD3, CAMP, LAMP, DRAMP and dbAMP	3268	3268	No	No	No	34, 22
MPMABP	Li et al. (2022b)	CNN and BiLSTM	MLBP	2409	–	Yes	Yes	No	7, 6
TPpred-ATMV	Yan et al. (2022)	AMVTLF	AntiBP and AVPpred	–	–	No	Yes	No	29, 26
AMP-GSM	(Gülsüm Söylemez et al., 2023)	RF	DBAASP, APD vr.3 and AIPpred	1258	1887	No	No	No	2, 2
CAMPR4	(Gawde et al., 2023)|	RF, SVM or ANN	NCBI protein, PDB, PubMed	3920	3920	No	No	Inaccessible	45, 27

Algorithm names: random forest, RF; support vector machine, SVM; convolutional neural networks, CNN; bidirectional long short term memory, Bi-LSTM; adaptive multi-view based on the tensor learning framework, AMVTLF; adaptive neuro-fuzzy inference system, ANFIS; synthetic minority over-sampling technique, SMOTE; deep neural networks, DNN; k nearest neighbor, kNN.

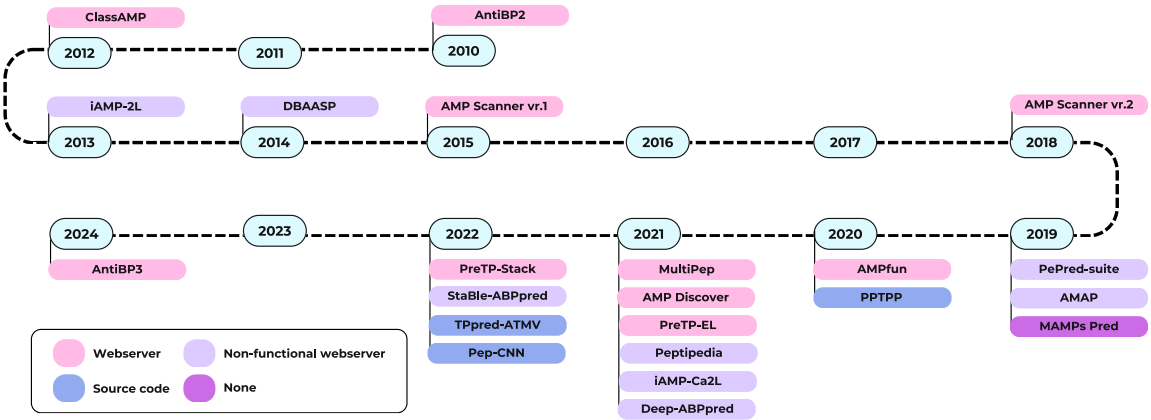

Fig. 3. Chronology of specific antimicrobial classifiers.

Model performance was assessed using receiver operating characteristic (ROC) curves, with the AUC serving as the performance metric. Additionally, classifier performance was evaluated using standard metrics, including sensitivity, specificity, and error rate.[3] As these metrics require the definition of a threshold for classification, evaluations were conducted using both the default threshold (set by the classifier) and the optimal threshold. Classifiers PreAIP, AnOxPePred, AnOxPP and MultiPep output a score between 0 and 1, with the default being 0.5. PreTP-Stack and PreTP-EL, on the other hand, provide the predicted category, positive or negative, and a corresponding score for between 0 and 1 for the predicted category. To account for this, if the predicted class was negative, the score was modified to be negative so higher negative scores indicate stronger non-anti-inflammatory activity with 0 separating the two. Finally, MLBP (Tang et al., 2022) does not output a probability but merely a binary category, which was categorized as 0 or 1 with no threshold (which meant that assessment using ROC curves was not meaningful, so it was excluded for this classifier). The optimal threshold, that allows for maximum discriminatory power between positive and negative peptides, was determined as the value that maximizes the F1-score, providing a balanced measure between precision and sensitivity.

3.2. Anti-inflammatory

Anti-inflammatory classification validation was limited to PreAIP, PreTP-EL, PreTP-Stack, and MLBP. Results from independent test set I indicate that PreAIP, PreTP-EL, and PreTP-Stack all performed well, achieving AUC values over 0.75 (Fig. 4A). At the default threshold, PreAIP and PreTP-Stack demonstrated both sensitivity and specificity greater than 50%, with error rates slightly above 25% (Fig. 4B). As these three classifiers were trained on the same dataset, independent set I serves as a common benchmark for comparison (see Table 7). Of the three, PreAIP performed the best, with the highest AUC score of 0.833 and lowest error rate. Despite being trained on the same dataset, the classifiers use different algorithms, which accounts for their performance differences. PreAIP is an RF classifier, while PreTP-EL and PreTP-Stack are ensemble classifiers. PreTP-EL combines RF and SVM, whereas PreTP-Stack integrates SVM, RF, linear discriminant analysis (LDA), extreme gradient boosting (XGB), and adaptive majority voting (AMV). In contrast, MLBP performed no better than a random classifier, with an AUC of 0.505 and an error rate close to 50%.

However, classifier performance was much worse when tested with independent test set II. ROC curves showed that all tools had predictions only slightly better than random, with curves close to the random 45-degree line (Fig. 4C). PreTP-Stack and PreAIP had similar AUC values of approximately 0.570 followed by PreTP-EL with an AUC of 0.495. Despite the overall poor performance, PreAIP, PreTP-EL, and MLBP demonstrated relatively high specificity at their default thresholds, indicating a lower rate of false positives relative to true negatives (Fig. 4D). However, their low sensitivity indicates a high false negative rate, meaning bioactive peptides are more likely to be misclassified than non-bioactive peptides. This high false negative rate may result from the classifiers not being general enough to recognize all bioactive peptides, possibly due to the small size of the training database. To further investigate performance, we grouped peptides in the second independent test dataset based on peptide length. Given that most peptides used to train the classifiers were 15–25 amino acids long, we aimed to assess if peptide length impacted the results. Indeed, PreTP-Stack and MLBP displayed higher error rates with longer peptides, while the other classifiers displayed lower error rates (Fig. 4F), further highlighting the important of large training sets.

3.3. Antioxidant

Antioxidant classification validation was limited to AnOxPP, AnOxPePred (including both Scavenger and Chelator) and MultiPep. In contrast to anti-inflammatory classifiers, the results from independent test set I indicate that all classifiers failed to perform better than a random classifier, with ROC curves close to the 45° line (Fig. 5A). MultiPep with the highest AUC score of 0.600, has the best discriminatory power. While the sensitivity is low for AnOxPePred-Scavenger, AnOxPePred-Chelator and MultiPep, the specificity is high, being close to 100%, which indicates that the ratio of false positives to true

[3] Sensitivity is defined as the ratio of true positives to all positives; specificity is the ratio of true negatives to all negatives; error rate is the ratio of number of peptides classified incorrectly to total number of peptides in the test set.

Table 6
Summary of specific antimicrobial classifiers.

Name	Reference	Algorithm	Positive database	Positive peptides	Negative peptides	Data availability	Offline code	Webserver	Citations (all, articles)
AntiBP2	Lata et al. (2010)	SVM	APD	999	999	No	No	Yes	196, 139
ClassAMP	Joseph et al. (2012)	SVM or RF	CAMP	AVP(54), ABP(454), AFP(61)	AVP(108), ABP(908), AFP(122)	No	No	Yes	112, 81
iAMP-2L	Xiao et al. (2013)	kNN	APD	1486	2405	No	No	Inaccessible	441, 355
DBAASP	Gogoladze et al. (2014)	Physicochemical	PubMed	1083	–	No	No	Inaccessible	86, 49
Antimicrobial Peptide Scanner vr.1	Veltri (2015)	RF	APD	115	116	No	No	Yes	10,
Antimicrobial Peptide Scanner vr.2	Veltri et al. (2018)	DNN	APD3	712	712	No	No	Yes	265, 199
MAMPs Pred	Lin et al. (2019)	RF	APD	2618	4371	No	No	No	34, 22
AMAP	Gull et al. (2019)	SVM	APD3	2704	5156	No	No	Inaccessible	53, 34
PEPred-suite	Wei et al. (2019)	RF	AntiBP and AVPpred	1258	1887	No	No	Inaccessible	117, 98
AMPfun	Chung et al. (2020)	RF, SVM	APD3, ADAM, ParaPep, AVPdb, CancerPPD, MLACP, AntiCP, AntiFP, and DRAMP	AVP(1400), APP(140), ABP(1930, 1931), AFP (1912)	AVP(2451), APP(700), ABP(1624, 1634), AFP (1261)	Yes	No	Yes	0, 0
PPTPP	Zhang and Zou (2020)	RF	AntiBP and AVPpred	AVP(544), ABP(800)	AVP(405), ABP(800)	No	Yes	No	68, 59
iAMP-Ca2L	Xiao et al. (2021)	CNN-BiLSTM-SVM	APD3, AMPer, and ADAM	3594	3925	Yes	Yes	Inaccessible	36, 29
AMP Discover	Pinacho-Castellanos et al. (2021)	RF	starPepDB	9781	9767	No	No	Yes	25, 22
Deep-ABPpred	Sharma et al. (2021b)	BiLSTM	APD, DRAMP, and MilkAMP	1635	1485	Yes	No	Inaccessible	49, 42
MultiPep	Grønning et al. (2021)	CNN	APD3, BioDADPep, BIOPEP-UWM, CAMPR3, DBAASP, PeptideDB, SATPdb, etc	14362, (13538 ABP)	–	Yes	Yes	Yes	15, 11
Peptipedia	Quiroz et al. (2021)	RF	UniProt, LAMP2, SATPdb, DBAASP, DRAMP, CAMP, etc.	–	–	No	No	Inaccessible	14, 8
PreTP-EL	Guo et al. (2021)	Ensemble	AntiBP and AVPpred	AVP(544), ABP(800)	AVP(405), ABP(800)	Yes	No	Yes	28, 23
StaBle-ABPpred	Singh et al. (2022)	BiLSTM	APD, DRAMP, and MilkAMP	652	7284	Yes	No	Inaccessible	24, 19
Pep-CNN	Zhang and Li (2022)	CNN	AntiBP and AVPpred	AVP(544), ABP(800)	AVP(405), ABP(800)	Yes	Yes	No	8, 5
TPpred-ATMV	Yan et al. (2022)	AMVTLF	AntiBP and AVPpred	–	–	No	Yes	No	29, 26
PreTP-Stack	Yan et al. (2023)	Ensemble	AntiBP and AVPpred	AVP(544), ABP(800)	AVP(405), ABP(800)	Yes	No	Yes	13, 10
AntiBP3	Bajiya et al. (2024)	RF	APD3, AntiBP2, dbAMP 2.0, CAMPR3, DRAMP, and ABP-Finder	744 g positive, 1164 g negative, 1797 g variable	–	Yes	Yes	Yes	0, 0

Algorithm names: random forests, RF; support vector machine, SVM; convolutional neural networks, CNN; bidirectional long short term memory, Bi-LSTM; adaptive multi-view based on the tensor learning framework, AMVTLF; deep neural networks, DNN; k nearest neighbor, kNN.

Table 7
Number of peptides used to test each classifier.

Bioactivity category	Classifier name	Positive peptides	Negative peptides
Anti-inflammatory	MLBP	2867	515
Anti-inflammatory	PreTP-EL	1118	892
Anti-inflammatory	PreAIP	1118	892
Anti-inflammatory	PreTP-Stack	1118	892
Antioxidant	AnOxPP	136	46
Antioxidant	AnOxPePred	362	1057
Antioxidant	MultiPep	387	1103

negatives is low (Fig. 5B). AnOxPP did not perform as well as the others with sensitivity, specificity and error rate at 50%. Results from independent test set II were better (Fig. 5C–F). MultiPep was again found to have the highest AUC of 0.852 followed by AnOxPePred-Chelator with 0.602. At the default threshold, AnOxPePred-Scavenger, AnOxPePred-Chelator and MultiPep had a high specificity but low sensitivity, indicating that these classifiers have problems with false negatives. Lowering the threshold can lower the number of false negatives and increase sensitivity, but this comes at the cost of increasing the number of false positives, thereby decreasing specificity. Taken together, all models have an error rate just under 50% suggesting that they are only

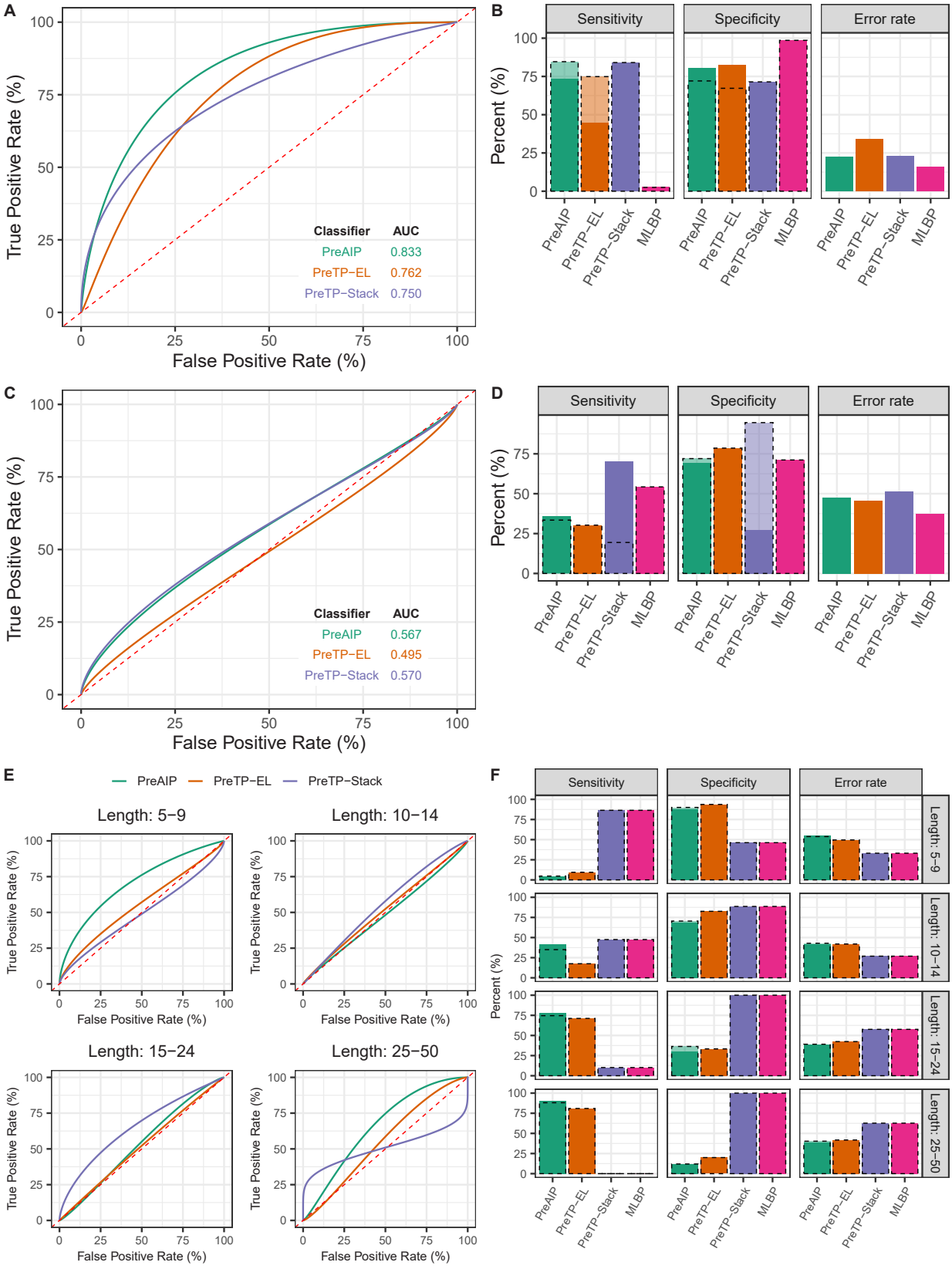

Fig. 4. Results from anti-inflammatory classifier evaluation. (A) ROC curves and the corresponding AUC values of classifiers evaluated with independent test set I. (B) sensitivity, specificity and error rate of classifiers evaluated with independent test set I. (C) ROC curves and the corresponding AUC values of classifiers evaluated with independent test set II. (D) sensitivity, specificity and error rate of classifiers evaluated with independent test set II. (E) ROC curves for peptides of classifiers grouped by peptide length of independent test set II. (F) sensitivity, specificity and error rate of classifiers grouped by peptide length of independent test set II. Note: for the bar plots, dotted lines indicate the optimal threshold, whereas solid lines represent the default threshold.

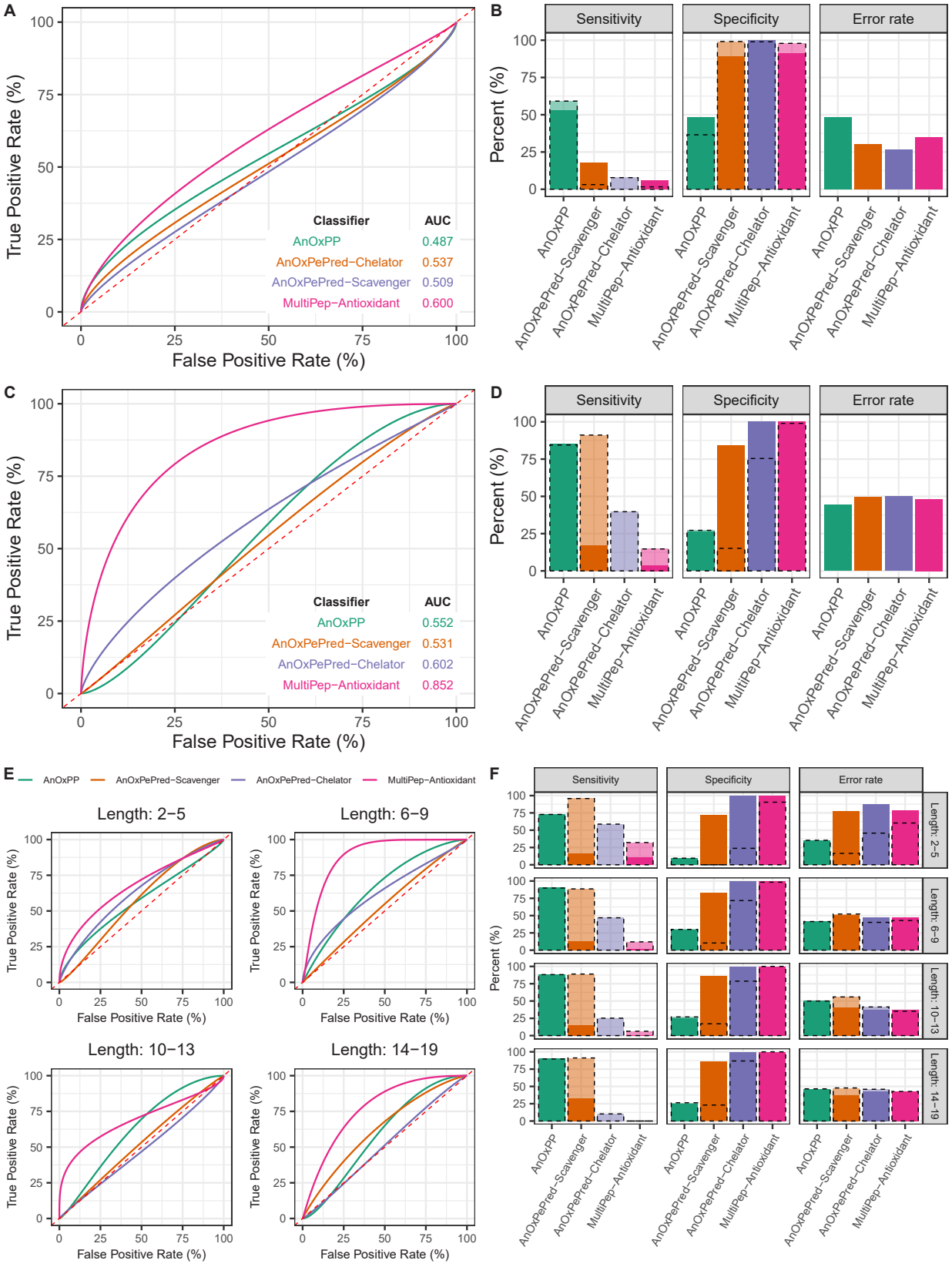

Fig. 5. Results from antioxidant classifier evaluation. (A) ROC curves and the corresponding AUC values of classifiers evaluated with independent test set I. (B) sensitivity, specificity and error rate of classifiers evaluated with independent test set I. (C) ROC curves and the corresponding AUC values of classifiers evaluated with independent test set II. (D) sensitivity, specificity and error rate of classifiers evaluated with independent test set II. (E) ROC curves for peptides of classifiers grouped by peptide length of independent test set II. (F) sensitivity, specificity and error rate of classifiers grouped by peptide length of independent test set II. Note: for the bar plots, dotted lines indicate the optimal threshold, whereas solid lines represent the default threshold.

slightly better than a random classifier. No major differences in results were observed when the peptide validation set was split by peptide length, with the exception of error rate, which was much higher for small peptides with 2–5 residues.

4. Conclusions

The peptides that make up hydrolysate products are more than just the sum of their amino acids, and peptide bioactivity classification serves as an important (if small) step towards predicting the impact that any given hydrolysate product will have in cell culture. Accurate classification of, for example, anti-oxidant or anti-microbial activity in hydrolysate products can not only guide product selection, but lead to the development of new and more potent hydrolysate formulations. Despite the large number of published classifiers, however, serious gaps remain in the application of bioactivity prediction in media optimization. First and foremost, the vast majority of published classifiers have focused on AMP prediction, and relatively few have been developed for predicting antioxidant peptides, growth factor peptides, or peptide hormones. A greater emphasis on the latter categories will be essential for media optimization applications. Second is the issue of reproducibility. Although classifiers available as a webserver garnered significantly higher research interest, webserver maintenance remains a challenge, with almost half of published servers currently non-functional. Among the tools offered offline, many could not be used as a result of bugs or insufficient documentation. A greater focus on accessibility and usability will no doubt aid the adoption of bioactivity classification tools. Third is the issue of generalizability. Despite relatively good performance on test sets similar to training sets, the evaluation of anti-inflammatory and antioxidant classifiers using independent test sets revealed that current models generally suffer from low accuracy, with AUC scores and error rates close to 50%. This suggests that existing ML based approaches cannot be solely relied upon to identify novel peptides. Moreover, the low sensitivity and specificity observed by many classifiers indicate that these tools are not effective for preliminary screening of peptides, as many potential bioactive peptides will likely be lost in the process and non-bioactive peptides retained. Some of this poor performance is likely a consequence of the limited size of the training sets used to construct the classifiers. Another cause is the lack of reliable databases for peptides confirmed to be non-bioactive, with peptides assumed to be non-bioactive later identified to have bioactive properties. Despite the challenges, none of the issues identified in this review are insurmountable — continued interest in peptide bioactivity will no doubt lead to more standardized tools and the development of comprehensive peptide repositories, and with them, the promise of more rational media design.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All independent test sets are available at https://doi.org/10.5281/zenodo.11402692.

Machine Learning Tools for Peptide Bioactivity Evaluation Implications for Cell Culture Media Optimization and the Broader Cultivated Meat Industry (Original data) (Zenodo)

Acknowledgments

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2019-04694] and The Good Food Institute, Inc. under GFI grant number 21-CM-CA-EG-1-04. KSI was partially funded by the Dalhousie University Master's Vitamin Scholarship, the Nova Scotia Graduate Scholarship (Master's), and the Dr. Robert Gillespie Scholarship. MC was partially funded by the Dalhousie University Doctoral Vitamin Scholarship.

References

Ahmed, S., Muhammod, R., Khan, Z.H., Adilina, S., Sharma, A., Shatabda, S., Dehzangi, A., 2021. ACP-MHCNN: an accurate multi-headed deep-convolutional neural network to predict anticancer peptides, 12 Sci. Rep. 11.

Alotaibi, F., Attique, M., Khan, Y.D., 2021. AntiFlamPred: an anti-inflammatory peptide predictor for drug selection strategies. Comput. Mater. Continua (CMC) 69, 1039–1055.

Bajiya, N., Choudhury, S., Dhall, A., Raghava, G.P., 2024. AntiBP3: a method for predicting antibacterial peptides against gram-positive/negative/variable bacteria, 2 Antibiotics 13.

Bárcenas, O., Pintado-Grima, C., Sidorczuk, K., Teufel, F., Nielsen, H., Ventura, S., Burdukiewicz, M., 2022. The dynamic landscape of peptide activity prediction, 1 Comput. Struct. Biotechnol. J. 20, 6526–6533.

Bejani, M.M., Ghatee, M., 2021. A systematic review on overfitting control in shallow and deep neural networks, 12 Artif. Intell. Rev. 54, 6391–6438.

Benfield, A.H., Henriques, S.T., 2020. Mode-of-action of antimicrobial peptides: membrane disruption vs. intracellular mechanisms. Front Med. Technol. 2.

Bhadra, P., Yan, J., Li, J., Fong, S., Siu, S.W., 2018. AmPEP: sequence- based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest, 12 Sci. Rep. 8.

Burdukiewicz, M., Sidorczuk, K., Rafacz, D., Pietluch, F., Chilimoniuk, J., Rödiger, S., Gagat, P., 2020. Proteomic screening for prediction and design of antimicrobial peptides with ampgram, 6 Int. J. Mol. Sci. 21, 1–13.

Carus, M., Scharf, A., Breitmayer, E., 2019. Review and gap-analysis of LCA-studies of cultured meat for the good food Institute. www.nova-institut.eu. (Accessed 28 May 2024).

Chandra, V., Tiwari, A., Pant, K.K., Bhatt, R., 2022. Animal cell culture: basics and applications, 1 Industrial Microbiology and Biotechnology 691–719.

Chen, L., Hu, Z., Rong, Y., Lou, B., 2024. Deep2Pep: a deep learning method in multi-label classification of bioactive peptide, 4 Comput. Biol. Chem. 109.

Chung, C.R., Kuo, T.R., Wu, L.C., Lee, T.Y., Horng, J.T., 2020. Characterization and identification of antimicrobial peptides with different functional activities, 5 Briefings Bioinf. 21, 1098–1114.

Cruz-Chamorro, I., Santos-Sánchez, G., Bollati, C., Bartolomei, M., Li, J., Arnoldi, A., Lammi, C., 2022. Hempseed (Cannabis sativa) peptides WVSPLAGRT and IGFLIIWV exert anti-inflammatory activity in the LPS-stimulated human hepatic cell line, 1 J. Agric. Food Chem. 70, 577–583.

Dayem, A.A., Lee, S.B., Lim, K.M., Kim, A., Shin, H.J., Vellingiri, B., Kim, Y.B., Cho, S.G., 2023. Bioactive peptides for boosting stem cell culture platform: methods and applications, 4 Biomed. Pharmacother. 160.

de Castro, R.J., Sato, H.H., 2014. Antioxidant activities and functional properties of soy protein isolate hydrolysates obtained using microbial proteases, 2 Int. J. Food Sci. Technol. 49, 317–328.

de Oliveira, E.C.L., Santana, K., Josino, L., e Lima, A.H.L., de Souza de Sales Júnior, C., 2021. Predicting cell-penetrating peptides using machine learning algorithms and navigating in their chemical space, 12 Sci. Rep. 11.

Deng, H., Lou, C., Wu, Z., Li, W., Liu, G., Tang, Y., 2022. Prediction of anti-inflammatory peptides by a sequence-based stacking ensemble model named AIPStack, 9 iScience 25.

Deng, W., Chen, Y., Sun, X., Wang, L., 2023. AODB: a comprehensive database for antioxidants including small molecules, peptides and proteins, 8 Food Chem. 418.

Djemal, L., von Hagen, J., Kolmar, H., Deparis, V., 2021. Characterization of soy protein hydrolysates and influence of its iron content on monoclonal antibody production by a murine hybridoma cell line, 7 Biotechnol. Prog. 37, e3147.

Farias, T.C., Abreu, J.P., Oliveira, J.P., Macedo, A.F., Rodríguez-Vega, A., Tonin, A.P., Cardoso, F.S., Meurer, E.C., Koblitz, M.G., 2023. Bioactive properties of peptide fractions from Brazilian soy protein hydrolysates: in silico evaluation and experimental evidence, 12 Food Hydrocolloids Health 3, 100112.

Fernandes, F.C., Rigden, D.J., Franco, O.L., 2012. Prediction of antimicrobial peptides based on the adaptive neuro-fuzzy inference system application. Biopolymers 98, 280–287.

Fingerhut, L.C., Miller, D.J., Strugnell, J.M., Daly, N.L., Cooke, I.R., 2020. Ampir: an R package for fast genome-wide prediction of antimicrobial peptides, 11 Bioinformatics 36, 5262–5263.

Fu, H., Cao, Z., Li, M., Wang, S., 2020. ACEP: improving antimicrobial peptides recognition through automatic feature fusion and amino acid embedding, 8 BMC Genom. 21.
Gaffar, S., Hassan, M.T., Tayara, H., Chong, K.T., 2024. IF-AIP: a machine learning method for the identification of anti-inflammatory peptides using multi-feature fusion strategy, 1 Comput. Biol. Med. 168.
Gawde, U., Chakraborty, S., Waghu, F.H., Barai, R.S., Khanderkar, A., Indraguru, R., Shirsat, T., Idicula-Thomas, S., 2023. CAMPR4: a database of natural and synthetic antimicrobial peptides, 1 Nucleic Acids Res. 51, D377–D383.
Gogoladze, G., Grigolava, M., Vishnepolsky, B., Chubinidze, M., Duroux, P., Lefranc, M. P., Pirtskhalava, M., 2014. DBAASP: database of antimicrobial activity and structure of peptides. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett. 357, 63–68.
Grønning, A.G., Kacprowski, T., Schéele, C., 2021. MultiPep: a hierarchical deep learning approach for multi-label classification of peptide bioactivities. Biology Meth. Protocols 6.
Gull, S., Minhas, F., 2022. AMP0: species-specific prediction of anti-microbial peptides using zero and few shot learning. IEEE ACM Trans. Comput. Biol. Bioinf 19, 275–283.
Gull, S., Shamim, N., Minhas, F., 2019. AMAP: hierarchical multi-label prediction of biologically active and antimicrobial peptides, 4 Comput. Biol. Med. 107, 172–181.
Gülsüm Söylemez, Ümmü, Yousef, M., Bakir-Gungor, B., 2023. AMP-GSM: prediction of antimicrobial peptides via a grouping-scoring-modeling approach, 4 Appl. Sci. 13.
Guo, Y., Yan, K., Lv, H., Liu, B., 2021. PreTP-EL: prediction of therapeutic peptides based on ensemble learning, 11 Briefings Bioinf. 22.
Gupta, S., Sharma, A.K., Shastri, V., Madhu, M.K., Sharma, V.K., 2017. Prediction of anti-inflammatory proteins/peptides: an insilico approach, 1 J. Transl. Med. 15.
Halliwell, B., 2014. Cell culture, oxidative stress, and antioxidants: avoiding pitfalls. Biomed. J. 37, 99–105.
Ho, Y.Y., Lu, H.K., Lim, Z.F.S., Lim, H.W., Ho, Y.S., Ng, S.K., 2021. Applications and analysis of hydrolysates in animal cell culture, 9 2021 Bioresour. Bioprocess. 8 (1 8), 1–15.
Holton, T.A., Pollastri, G., Shields, D.C., Mooney, C., 2013. CPPpred: prediction of cell penetrating peptides, 12 Bioinformatics 29, 3094–3096.
Hsieh, L.S., Lu, M.S., Chiang, W.D., 2022. Identification and characterization of immunomodulatory peptides from pepsin-soy protein hydrolysates, 12 Bioresour. Bioprocess. 9, 1–13.
Humbird, D., 2021. Scale-up economics for cultured meat, 8 Biotechnol. Bioeng. 118, 3239–3250.
Hussain, W., 2022. sAMP-PFPDeep: improving accuracy of short antimicrobial peptides prediction using three different sequence encodings and deep neural networks, 1 Briefings Bioinf. 23.
Joseph, S., Karnik, S., Nilawe, P., Jayaraman, V.K., Idicula-Thomas, S., 2012. ClassAMP: a prediction tool for classification of antimicrobial peptides. IEEE ACM Trans. Comput. Biol. Bioinf 9.
Kaur, D., Patiyal, S., Arora, C., Singh, R., Lodhi, G., Raghava, G.P., 2021. In-silico tool for predicting, scanning, and designing defensins, 11 Front. Immunol. 12.
Kavousi, K., Bagheri, M., Behrouzi, S., Vafadar, S., Atanaki, F.F., Lotfabadi, B.T., Ariaeenejad, S., Shockravi, A., Moosavi-Movahedi, A.A., 2020. IAMPE: NMR-assisted computational prediction of antimicrobial peptides, 10 J. Chem. Inf. Model. 60, 4691–4701.
Khatun, M.S., Hasan, M.M., Kurata, H., 2019. PreAIP: computational prediction of anti-inflammatory peptides by integrating multiple complementary features. Front. Genet. 10.
Kumar, M., Hasan, M., Choyal, P., Tomar, M., Gupta, O.P., Sasi, M., Changan, S., Lorenzo, J.M., Singh, S., Sampathrajan, V., Dhumal, S., Pandiselvam, R., Sharma, K., Satankar, V., Waghmare, R., Senapathy, M., Sayed, A.A., Radha Dey, A., Amarowicz, R., Kennedy, J.F., 2022. Cottonseed feedstock as a source of plant-based protein and bioactive peptides: evidence based on biofunctionalities and industrial applications, 10 Food Hydrocolloids 131, 107776.
Kwon, D.H., Cha, H.J., Lee, H., Hong, S.H., Park, C., Park, S.H., Kim, G.Y., Kim, S., Kim, H.S., Hwang, H.J., Choi, Y.H., 2019. Protective effect of glutathione against oxidative stress-induced cytotoxicity in RAW 264.7 macrophages through activating the nuclear factor erythroid 2-related factor-2/heme oxygenase-1 pathway, 4 Antioxidants 8.
Lata, S., Mishra, N.K., Raghava, G.P., 2010. AntiBP2: improved version of antibacterial peptide prediction, 1 BMC Bioinf. 11.
Lawrence, T.J., Carper, D.L., Spangler, M.K., Carrell, A.A., Rush, T.A., Minter, S.J., Weston, D.J., Labbé, J.L., 2021. AmPEPpy 1.0: a portable and accurate antimicrobial peptide prediction tool, 7 Bioinformatics 37, 2058–2060.
Lee, H.A., Kim, I.H., Nam, T.J., 2015. Bioactive peptide from Pyropia yezoensis and its anti-inflammatory activities, 12 Int. J. Mol. Med. 36, 1701–1706.
Lertampaiporn, S., Vorapreeda, T., Hongsthong, A., Thammarongtham, C., 2021. Ensemble-AMPPred: robust AMP prediction and recognition using the ensemble learning method with a new hybrid feature for differentiating AMPs, 2 Genes 12, 1–17.
Li, C., Sutherland, D., Hammond, S.A., Yang, C., Taho, F., Bergman, L., Houston, S., Warren, R.L., Wong, T., Hoang, L.M., Cameron, C.E., Helbing, C.C., Birol, I., 2022a. AMPlify: attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens, 12 BMC Genom. 23.
Li, Y., Li, X., Liu, Y., Yao, Y., Huang, G., 2022b. MPMABP: a CNN and Bi-LSTM-based method for predicting multi-activities of bioactive peptides, 6 Pharmaceuticals 15.
Lin, W., Xu, D., 2016. Imbalanced multi-label learning for identifying antimicrobial peptides and their functional types, 12 Bioinformatics 32, 3745–3752.
Lin, Y., Cai, Y., Liu, J., Lin, C., Liu, X., 2019. An advanced approach to identify antimicrobial peptides and their function types for penaeus through machine learning strategies, 6 BMC Bioinf. 20.
Lin, T.-T., Yang, L.-Y., Lu, I.-H., Cheng, W.-C., Hsu, Z.-R., Chen, S.-H., Lin, C.-Y., 2021. AI4AMP: an antimicrobial peptide predictor using physicochemical property-based encoding method and deep learning. mSystems 6.
Lv, H., Yan, K., Guo, Y., Zou, Q., Hesham, A.E.L., Liu, B., 2022. AMPpred-EL: an effective antimicrobial peptide prediction model based on ensemble learning, 7 Comput. Biol. Med. 146.
Mahmood, A., Ali, S., 2017. Microbial and viral contamination of animal and stem cell cultures: common contaminants, detection and elimination, 5 J. Stem Cell Res. Therapeut. 2.
Manavalan, B., Shin, T.H., Kim, M.O., Lee, G., 2018. AIPpred: sequence-based prediction of anti-inflammatory peptides using random forest, 3 Front. Pharmacol. 9.
Manavalan, B., Basith, S., Shin, T.H., Wei, L., Lee, G., 2019. MAHTPred: a sequence-based meta-predictor for improving the prediction of anti-hypertensive peptides using effective feature representation, 8 Bioinformatics 35, 2757–2765.
Manavalan, B., Basith, S., Lee, G., 2022. Comparative analysis of machine learning-based approaches for identifying therapeutic peptides targeting SARS-CoV-2, 1 Briefings Bioinf. 23.
Meher, P.K., Sahu, T.K., Saini, V., Rao, A.R., 2017. Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC, 2 Sci. Rep. 7.
Minkiewicz, P., Iwaniak, A., Darewicz, M., 2019. BIOPEP-UWM database of bioactive peptides: current opportunities, 12 Int. J. Mol. Sci. 20.
Mooney, C., Haslam, N.J., Pollastri, G., Shields, D.C., 2012. Towards the improved discovery and design of functional peptides: common features of diverse classes permit generalized prediction of bioactivity, 10 PLoS One 7.
Obaidi, I., Mota, L.M., Quigley, A., Butler, M., 2021. The role of protein hydrolysates in prolonging viability and enhancing antibody production of CHO cells, 4 Appl. Microbiol. Biotechnol. 105, 3115–3129.
Olsen, T.H., Yesiltas, B., Marin, F.I., Pertseva, M., García-Moreno, P.J., Gregersen, S., Overgaard, M.T., Jacobsen, C., Lund, O., Hansen, E.B., Marcatili, P., 2020. AnOxPePred: using deep learning for the prediction of antioxidative properties of peptides, 12 Sci. Rep. 10.
O'Neill, E.N., Cosenza, Z.A., Baar, K., Block, D.E., 2021. Considerations for the development of cost-effective cell culture media for cultivated meat production, 1 Compr. Rev. Food Sci. Food Saf. 20, 686–709.
Pandey, P., Patel, V., George, N.V., Mallajosyula, S.S., 2018. KELM-CPPpred: kernel extreme learning machine based prediction model for cell-penetrating peptides, 9 J. Proteome Res. 17, 3214–3222.
Pinacho-Castellanos, S.A., García-Jacas, C.R., Gilson, M.K., Brizuela, C.A., 2021. Alignment-free antimicrobial peptide predictors: improving performance by a thorough analysis of the largest available data set, 6 J. Chem. Inf. Model. 61, 3141–3157.
Pirtskhalava, M., Amstrong, A.A., Grigolava, M., Chubinidze, M., Alimbarashvili, E., Vishnepolsky, B., Gabrielian, A., Rosenthal, A., Hurt, D.E., Tartakovsky, M., 2021. DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics, 1 Nucleic Acids Res. 49, D288–D297.
Porto, W.F., Pires, Állan S., Franco, O.L., 2012. CS-AMPPred: an updated SVM model for antimicrobial activity prediction in cysteine-stabilized peptides, 12 PLoS One 7.
Qin, D., Bo, W., Zheng, X., Hao, Y., Li, B., Zheng, J., Liang, G., 2022. DFBP: a comprehensive database of food-derived bioactive peptides for peptidomics research, 6 Bioinformatics 38, 3275–3280.
Qin, D., Jiao, L., Wang, R., Zhao, Y., Hao, Y., Liang, G., 2023. Prediction of antioxidant peptides using a quantitative structure-activity relationship predictor (AnOxPP) based on bidirectional long short-term memory neural network and interpretable amino acid descriptors, 3 Comput. Biol. Med. 154.
Quiroz, C., Saavedra, Y.B., Armijo-Galdames, B., Amado-Hinojosa, J., Álvaro, Olivera-Nappa, Sanchez-Daza, A., Medina-Ortiz, D., 2021. Peptipedia: a user-friendly web application and a comprehensive database for peptide research supported by Machine Learning approach. Database 2021.
Santos-Júnior, C.D., Pan, S., Zhao, X.M., Coelho, L.P., 2020. Macrel: antimicrobial peptide screening in genomes and metagenomes, 12 PeerJ 8.
Saravanan, V., Lakshmi, P.T., 2015. ACPP: a web server for prediction and design of anti-cancer peptides. Int. J. Pept. Res. Therapeut. 21, 99–106.
Sarker, I.H., 2021. Machine learning: algorithms, real-world applications and research directions, 5 SN Computer Sci. 2.
Schaduangrat, N., Nantasenamat, C., Prachayasittikul, V., Shoombuatong, W., 2019. ACPred: a computational tool for the prediction and analysis of anticancer peptides. Molecules 24.
Sharma, R., Shrivastava, S., Singh, S.K., Kumar, A., Saxena, S., Singh, R.K., 2021a. AniAMPpred: artificial intelligence guided discovery of novel antimicrobial peptides in animal kingdom, 11 Briefings Bioinf. 22.
Sharma, R., Shrivastava, S., Singh, S.K., Kumar, A., Saxena, S., Singh, R.K., 2021b. Deep-Abppred: identifying antibacterial peptides in protein sequences using bidirectional LSTM with word2vec, 9 Briefings Bioinf. 22.
Sidorczuk, K., Gagat, P., Pietluch, F., Kala, J., Rafacz, D., Bakala, L., Slowik, J., Kolenda, R., Rödiger, S., Fingerhut, L.C., Cooke, I.R., MacKiewicz, P., Burdukiewicz, M., 2022. Benchmarks in antimicrobial peptide prediction are biased due to the selection of negative data. Briefings Bioinf. 23, 9.
Siemensma, A., Babcock, J., Wilcox, C., Huttinga, H., 2010. Towards an understanding of how protein hydrolysates stimulate more efficient biosynthesis in cultured cells. Protein Hydrolysat. Biotechnol. 33–54.
Singh, O., Hsu, W.L., Su, E.C.Y., 2021. Co-AMPpred for in silico-aided predictions of antimicrobial peptides by integrating composition-based features, 12 BMC Bioinf. 22.

Singh, V., Shrivastava, S., Singh, S.K., Kumar, A., Saxena, S., 2022. StaBle-ABPpred: a stacked ensemble predictor based on biLSTM and attention mechanism for accelerated discovery of antibacterial peptides, 1 Briefings Bioinf. 23.
Spearman, M., Lodewyks, C., Richmond, M., Butler, M., 2014. The bioactivity and fractionation of peptide hydrolysates in cultures of CHO cells, 5 Biotechnol. Prog. 30, 584–593.
Su, X., Xu, J., Yin, Y., Quan, X., Zhang, H., 2019. Antimicrobial peptide identification using multi-scale convolutional network, 12 BMC Bioinf. 20.
Su, R., Hu, J., Zou, Q., Manavalan, B., Wei, L., 2020. Empirical comparison and analysis of web-based cell-penetrating peptide prediction tools, 3 Briefings Bioinf. 21, 408–420.
Sun, X., Wang, K., Gao, S., Hong, H., Zhang, L., Liu, H., Feng, L., Luo, Y., 2021. Purification and characterization of antioxidant peptides from yak (Bos grunniens) bone hydrolysates and evaluation of cellular antioxidant activity, 8 J. Food Sci. Technol. 58, 3106–3119.
Tang, W., Dai, R., Yan, W., Zhang, W., Bin, Y., Xia, E., Xia, J., 2022. Identifying multi-functional bioactive peptide functions using multi-label deep learning, 1 Briefings Bioinf. 23.
Tuomisto, H.L., Mattos, M.J.T.D., 2011. Environmental impacts of cultured meat production, 7 Environ. Sci. Technol. 45, 6117–6123.
Ümmü, Gülsüm Söylemez, Yousef, M., Kesmen, Z., Büyükkiraz, M.E., Bakir-Gungor, B., 2022. Prediction of linear cationic antimicrobial peptides active against gram-negative and gram-positive bacteria based on machine learning models, 4 Appl. Sci. 12.
Veltri, D.P., 2015. A Computational and Statistical Framework for Screening Novel Antimicrobial Peptides. George Mason University (Ph.D. thesis phD thesis).
Veltri, D., Kamath, U., Shehu, A., 2018. Deep learning improves antimicrobial peptide recognition, 8 Bioinformatics 34, 2740–2747.
Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D. K., Sette, A., Peters, B., 2019. The immune epitope database (IEDB): 2018 update. Nucleic Acids Res. 47, D339–D343.
Wang, P., Hu, L., Liu, G., Jiang, N., Chen, X., Xu, J., Zheng, W., Li, L., Tan, M., Chen, Z., Song, H., Cai, Y.D., Chou, K.C., 2011. Prediction of antimicrobial peptides based on sequence alignment and feature selection methods. PLoS One 6.
Wei, L., Zhou, C., Su, R., Zou, Q., 2019. PEPred-Suite: improved and robust prediction of therapeutic peptides using adaptive feature representation learning, 11 Bioinformatics 35, 4272–4280.
Xiao, X., Wang, P., Lin, W.Z., Jia, J.H., Chou, K.C., 2013. iAMP-2L: a two-level multi-label classifier for identifying antimicrobial peptides and their functional types, 5 Anal. Biochem. 436, 168–177.
Xiao, X., Shao, Y.T., Cheng, X., Stamatovic, B., 2021. iAMP-CA2L: a new CNN-BiLSTM-SVM classifier based on cellular automata image for identifying antimicrobial peptides and their functional types, 11 Briefings Bioinf. 22.
Xu, J., Li, F., Leier, A., Xiang, D., Shen, H.H., Lago, T.T.M., Li, J., Yu, D.J., Song, J., 2021. Comprehensive assessment of machine learning-based methods for predicting antimicrobial peptides, 9 Briefings Bioinf. 22.
Yan, J., Bhadra, P., Li, A., Sethiya, P., Qin, L., Tai, H.K., Wong, K.H., Siu, S.W., 2020. Deep-AmPEP30: improve short antimicrobial peptides prediction with deep learning, 6 Mol. Ther. Nucleic Acids 20, 882–894.
Yan, K., Lv, H., Guo, Y., Chen, Y., Wu, H., Liu, B., 2022. TPpred-ATMV: therapeutic peptide prediction by adaptive multi-view tensor learning model, 5 Bioinformatics 38, 2712–2718.
Yan, K., Lv, H., Wen, J., Guo, Y., Xu, Y., Liu, B., 2023. PreTP-stack: prediction of therapeutic peptides based on the stacked ensemble learing, 3 IEEE ACM Trans. Comput. Biol. Bioinf 20, 1337–1344.
Yang, Y., Wu, H., Gao, Y., Tong, W., Li, K., 2023. MFPPDB: a comprehensive multi-functional plant peptide database. Front. Plant Sci. 14.
Zhang, S., Li, X., 2022. Pep-CNN: an improved convolutional neural network for predicting therapeutic peptides, 2 Chemometr. Intell. Lab. Syst. 221.
Zhang, Y.P., Zou, Q., 2020. PPTPP: a novel therapeutic peptide prediction method using physicochemical property encoding and adaptive feature representation learning, 7 Bioinformatics 36, 3982–3987.
Zhang, J., Zhang, Z., Pu, L., Tang, J., Guo, F., 2021a. AIEpred: an ensemble predictive model of classifier chain to identify anti-inflammatory peptides. IEEE ACM Trans. Comput. Biol. Bioinf 18, 1831–1840.
Zhang, Q.Y., Yan, Z.B., Meng, Y.M., Hong, X.Y., Shao, G., Ma, J.J., Cheng, X.R., Liu, J., Kang, J., Fu, C.Y., 2021b. Antimicrobial peptides: mechanism of action, activity and clinical potential, 12 Military Med. Res. 8.
Zhang, Y.Y., Stockmann, R., Ng, K., Broadbent, J.A., Stockwell, S., Suleria, H., Shaik, N.E. K., Unnithan, R.R., Ajlouni, S., 2023. Characterization of Fe(III)-binding peptides from pea protein hydrolysates targeting enhanced iron bioavailability, 3 Food Chem. 405, 134887.