Compilation and Mining
of
Peptide Hormones and their Receptors

Submitted by

Dashleen Kaur (PhD20209)

Under the guidance of
Prof. G.P.S Raghava
Head and Professor

In partial fulfilment of the requirements for the degree of
Master of Technology in Computational Biology

To

Department of Computational Biology,
Indraprastha Institute of Information Technology,
New Delhi

December 2022

Certificate

This is to certify that the thesis titled "**Compilation and Mining of Peptide Hormones and their Receptors**" being submitted by Dashleen Kaur to the Indraprastha Institute of Information Technology Delhi, for the award of the Master of Technology, is an original research work carried out by her under my supervision.

The results in this thesis have not been submitted in part or full to any other university or institute for awarding any degree/diploma.

April 2023

Prof. G.P.S Raghava

Head & Professor (Department of Computational Biology)

Indraprastha Institute of Information Technology Delhi- 110020

Acknowledgement

I would like to express my sincere gratitude and respect towards Prof. Gajendra P.S. Raghava from Indraprastha Institute of Information Technology, Delhi for being my supervisor and for exposing me to this wonderful topic of research and guiding me throughout. Besides my supervisor. I would like to thank the PhD scholar Akanksha Arora, Dr Sumeet, Dr Neelam and Research Associate Dr Naorem Leimarembi Devi from his esteemed lab for their constant guidance, support, and motivation throughout the project. I would also like to thank the Department of Computational Biology and IT administrators at IIIT Delhi for providing me with all the necessary resources. Lastly, I would also like to thank my family and friends for providing much-needed support and motivating me from time to time throughout the course of my thesis, which enabled me to pursue my research in an efficient and structured manner.

Dashleen Kaur

M.Tech CB (PhD20209)

Table of Contents

List of abbreviations

AAC	Amino acid composition
Acc	Accuracy
AUC	Area Under the Receiver Operating Characteristic curve
BLAST	Basic Local Alignment Search Tool
Blastp	protein-protein Basic Local Alignment Search Tool
CD-HIT	Cluster Database at High Identity with Tolerance
CV	Cross Validation
DT	Decision Tree
E-value	Expect value
GNB	Gaussian Naive Bayes
HTML	HyperText Markup Language
IR	Insulin Receptor
KNN	k-nearest neighbors
GBM	Gradient Boosting Machine
LR	Logistic Regression
MAST	Motif Alignment & Search Tool
MCC	Matthews correlation coefficient
MERCI	Motif-EmeRging and with Classes-Identification
ML	Machine Learning
MLP	Multi-layer Perceptron
NCBI	National Center for Biotechnology Information
RFE	Recursive Feature Elimination
RF	Random Forest
Sens	Sensitivity
Sklearn	Scikit-learn
Spec	Specificity
SVC	Support Vector Classifier
UniProtKB	UniProt Knowledgebase
VF	Virulence Factor

XGB	Extreme Gradient Boosting

List of Figures

List of Tables

Abstract

Hormones play a crucial role in communicating information between cells and organs; responsible for regulating almost all the physiological processes of organisms. Thus, it is important to collect, compile and mine hormones associated information. Firstly, a repository Hmrbase2 have been developed to maintain comprehensive information on hormones and their receptors, which is an update of Hmrbase. The information was compiled from literature and public repositories like HMDB, Uniprot, HORDB, ENDONET and PubChem. It contains a total of 12,056 entries, including 7,406 entries for peptide hormones, 753 entries for non-peptide hormones, and 3,897 entries for hormone receptors. The database also includes 5,662 hormone receptor pairs. The database is available free for scientific community (https://webs.iiitd.edu.in/raghava/hmrbase2/. Secondly, systematic attempt has been made to develop a method for predicting peptide hormones using data mining techniques. All models in this study were trained, test and evaluated on a dataset of 1174 hormonal and 1174 non-hormonal peptides. A wide range of machine and deep learning techniques have been implemented to discriminate hormones and non-hormones with high precision. Best performing model based on logistic regression achieved maximum performance AUC of 0.93. Finally, a hybrid method has been developed that combine logistic regression model (alignment free method) with BLAST/motif (alignment-based method) and achieved AUC of 0.96 with MCC of 0.8 on independent/validation dataset. To facilitate research community a web server HOPPred have been developed (https://webs.iiitd.edu.in/raghava/hoppred/).

Chapter-1

Introduction

The endocrine system is essential for the control and coordination of an organism. It regulates crucial life processes like growth, development and reproduction via complex signal transduction processes. The extensive cross-talk between various cells and organs for regulation is possible due to hormones. Hormones are the organic molecules produced in plants and animals for physiological regulation and homeostasis maintenance. The definition of hormone has changed since its original definition by Sterling in 1905, who defined it as a chemical secreted internally by glands that carry signals to target organs through the blood (1). Hormones could be produced by a specialised group of cells and could be transported via many means to their target cells. They bind to specific receptors in the target cells, culminating in desired changes in the target cells (2).

Hormones are messenger molecules created from different types of compounds found in the body, like proteins and lipids. The release of hormones into the bloodstream for endocrine coordination presents manufacturing, delivery, and response methods challenges. Hormones need receptors on the surface and nuclei of the cells to execute their effects effectively. Feedback loops regulate hormone actions, and endocrine disorders can occur due to overproduction, underproduction, autoimmune diseases, or genetic mutations. There are a number of endocrine conditions that can impact baby and toddler development, including diabetes, adrenal imbalance, thyroid illness, growth retardation, and Cushing's syndrome. The endocrine system also changes throughout life, with unique processes during pregnancy, puberty, and aging, which will be discussed.

There are several online resources available that provide information about important chemical molecules in our bodies. Presenting insights into hormones and their receptors and coverage on their type, source organism, sequence, and other chemical properties, is very limited. A single platform that provides information on hormones and their receptors can save time and energy for those seeking information and facilitate informed decision-making. Currently, no website offers a comprehensive and updated view of hormones and their receptors. Therefore, we have attempted to develop Hmrbase2.0, an updated version of Hmrbase, developed in 2009. Hmrbase2 is a database designed for endocrinologists, both theoretical and clinical, and contains information about hormones and their receptors. It is easy to use and includes information about the precursor protein sequences of mature hormone sequences, which can be used by researchers to study binding affinity and design better ligands for specific receptors. The database also includes information about protein domains, which are essential for understanding protein interactions.

Peptide drugs are not the traditional forms of therapeutics but are more specific and selective than conventional small molecule therapeutics. To improve the design and development of peptide therapeutics, computational approaches are used to understand the mechanism behind these drugs. These techniques can aid in discovering and developing new peptide therapeutics. Predicting peptide hormones is crucial for reducing cost and time for otherwise laborious and time-consuming drug discovery pipelines. Most studies work on the model for the prediction of other therapeutic peptides. In our study, we have worked on the model to predict whether the given peptide is hormonal or not. This study created models using a dataset of 5729 peptide hormones sequences. These models were tested and evaluated using various features to improve their accuracy. The resulting tool, called HopPred, can predict peptide hormones with a high level of precision accurately.

Chapter - 2

Compilation of peptide, non-peptide hormones and their receptors

INTRODUCTION

Hormones are essential for regulating various physiological processes, and hormone imbalances can contribute to the development of many health conditions, such as cancer, osteoporosis, and diabetes, making it necessary to gather and organise data about hormones and their receptors.

The chemical nature of hormones could be peptide or non-peptide. The peptide hormones are encoded by the genome (3). Upon synthesis, peptide hormones might undergo post-translational modifications and are generally packaged into secretory granules, secreted from the cells upon further stimuli. The spectrum of structures and lengths of peptide hormones across the plant and animal kingdoms can be attributed to their early rise in evolution. It could be exemplified by the similarity between gonadotropin-releasing hormone (GnRH) structures in mammals and the mating factor in yeast (4). Non-peptide hormones are derivatives of molecules and usually have identical structures across all organisms. Amino acid analogues are derived from amino acids, and many steroid hormones are derived from cholesterol. The thing that unifies hormones is their specific binding to the target cells. This is ensured by the presence of specific intracellular or cell-surface hormone receptors in the target cells. The peptide hormones usually bind to the extracellular receptors, whereas the non-peptide hormones bind to the intracellular receptors in the target cells (5). The cell-surface receptors like G-coupled and enzyme-linked receptors initiate an intracellular signal cascade upon binding to their respective ligands, which eventually activates cellular responses. The intracellular receptors belong to the “nuclear receptor superfamily” (6). The intracellular hormone-receptor complex controls the activities of responsive genes.

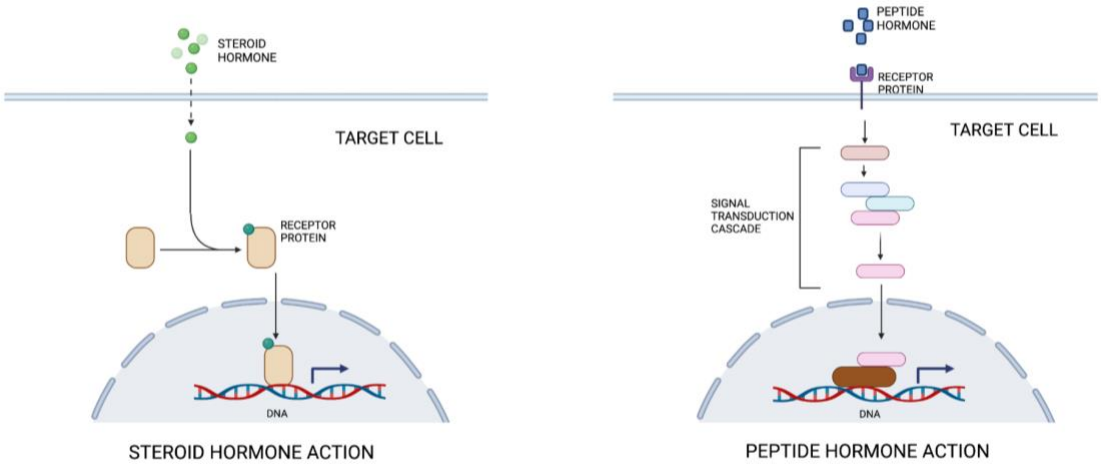

Figure 1: An overview of peptide and non-peptide hormones

Hormones play significant roles in controlling and regulating many physiological processes involving regulation of bioenergetics, reproduction, excretion and homeostasis. Hormone binding to the receptors alters the activity of the cells, and the increased or decreased cellular processes translate to regulated biological actions. Insulin is an example of a peptide hormone produced by the beta cells of pancreatic islets in vertebrates and many other animals (7). Like many other peptide hormones, it is also produced as a prohormone form which is inactive and can't act as a functional hormone (8). After proper cleavages, foldings and post-translational modifications, a functional mature hormone is packaged into granules (9). The actual effect of the hormone is initiated after it binds to its receptor in the target cell. The receptor of insulin is called the insulin receptor. Insulin receptor or IR is a receptor on the surface of cells belonging to the receptor tyrosine kinase family (10). Like many other surface receptors of the hydrophobic ligands, IR is also a transmembrane glycoprotein. Upon binding to the ligand, there is a conformation change in the receptor. This eventually leads to a cascade of recruitment and activation of proteins, which eventually culminate in the final intracellular effect (11). The physiological effect of insulin includes fat and glycogen synthesis, increased uptake of glucose and decreased lipolysis (12-14). Both excess and insufficient levels of hormones and defects in their reception by the target cells cause major life-altering endocrine disorders. Diabetes mellitus is a very well-known endocrine disease to exemplify this. Insufficient insulin production or type 1 diabetes is caused due to the demolishment of beta cells of the pancreas. This cause hyperglycemia or high blood sugar level, which interrupts the normal functioning of all major organ systems, including the cardiovascular and urinary system. A

decrease in the activity of insulin receptors can also cause insulin resistance, a cause of Type 2 diabetes (15). Endocrine disorders of low hormone levels include hypoglycemia and hypothyroidism (16).

Steroid hormones are hormones produced from cholesterol by the gonads, adrenal gland, and other tissues. Steroidogenesis, a process used to synthesise steroids, entails the formation of pregnenolone from cholestrol and then into certain steroid hormones. Steroid hormones are carried in the bloodstream bound to proteins and can also be stored in tissues like fat, muscle, and liver (17). Like in the case of peptide hormones, the binding of a steroid hormone to its specific receptor initiates a complex series of events within the cell, leading to changes in gene expression and the regulation of various physiological processes. However, since steroid hormones are lipophilic (fat-soluble) compounds that can pass through the cell membrane, they usually bind to intracellular receptors (18-19). Endocrine disorders involving steroid hormones involve an imbalance or dysfunction in the production, transport, or signalling of these hormones. These disorders can affect various aspects of health. They can have various symptoms, such as excess facial and body hair, acne, irregular periods, fertility problems, weight gain, high blood pressure, osteoporosis, fatigue, and cold intolerance. Examples of endocrine disorders involving steroid hormones include androgen excess disorders, estrogen excess disorders, progesterone deficiency disorders, glucocorticoid excess disorders, and thyroid hormone imbalances. These disorders can be caused by various factors, including genetic mutations, environmental stress, and pathogens or toxins. Treatment may involve medications, hormone replacement therapy, or lifestyle changes (20).

Phytohormones or hormones produced by plants regulate their growth, development, and response to environmental stimuli. There are several types of phytohormones, including auxins, cytokinins, gibberellins, abscisic acid, and brassinosteroids (21-22). Plant endocrine disorders can occur when there is an imbalance or dysfunction in phytohormone production, transport, or signalling. They can be caused by genetic mutations, environmental stress, and pathogens or toxins. These disorders can impact plant growth and overall health. Steroid phytohormones, derived from cholesterol and have a similar structure to animal steroids, play important roles in plant growth, development, and stress response. They include brassinosteroids, cytochromes P450, and flavonoids. Ongoing research on phytohormones is helping to improve our understanding of these regulatory mechanisms and how they can be manipulated to improve plant growth and productivity (23).

Several databases provide information on receptors and ligands, such as GPCRDB, BitterDB, and NR-DBIND (24-26). However, none of these resources offers comprehensive information on hormones and their receptors like Hmrbase, which was first developed in 2009. Hmrbase has helped to create other resources, such as SATPdb (27) and FeptideDB (28). As our understanding of endocrinology has grown, it has become necessary to update and improve the information available on hormones and their receptors. Hmrbase2 is an updated edition of this resource that provides detailed information on hormones and their receptors, including their localisation, functions, and sequence information. It also includes information on the domain and pharmaceutical use of these molecules, which was not included in the original version. Overall, Hmrbase2 is a comprehensive and user-friendly resource for information on hormones and their receptors.

MATERIALS AND METHODS

Collection of Data

Numerous pertinent information about peptide and non-peptide hormones, as well as their receptors, are dispersed in various sources like literature, databases, and other online sources. Information about hormones and hormone receptors was gathered by searching web resources and databases using keywords such as "hormone", "phytohormone", and "hormone receptors". Various web resources and databases, including HMDB (https://hmdb.ca/), Uniprot (https://www.uniprot.org/), HorDB (http://hordb.cpu-bioinfor.org/), Endonet (http://endonet.sybig.de/), Pubmed (https://pubmed.ncbi.nlm.nih.gov/) and PubChem (https://pubchem.ncbi.nlm.nih.gov/) were manually curated for the information.

Web-based interfaces and database architecture

The Hmrbase2 was created on a standard platform of Linux, Apache, MySQL and PHP (LAMP). Apache was used as the HTTP server, and MySQL was used to build the back end. To ensure the front end was accessible on various devices such as smartphones, tablets and desktops, HTML 5, CSS, Javascript and PHP 7.3.21 were used. The PHP programming language was then employed to create a single interface.

Data Content of the Database

This database hosts information about the hormones and hormone receptors in their sources, cellular, physical and functional properties and sequences. The fields representing these informations are specific for each type of hormones and receptors and include:

Peptide hormones - includes (i)'Organism' and (ii)'Taxonomy' of the source organism; (iii)Subcellular location: tells if the the hormone is secreted or the specific location of the hormone in the cell; (iv)Post translational modification: it gives the information if any other molecule has been attached to the mature peptide like phosphorylation; (v)Function: gives a brief function of the hormone; (vi)Sequence: give the amino acid sequence of mature peptide hormone; (vii)Receptor: The corresponding receptor to which the hormone binds; (viii)Uniprot ID

Non peptide hormones - include (i)Synonym: gives all the other names for the given hormone; (ii)Description; (iii)Formula: gives the scientific formula of the molecule; (iv)IUPAC: give the formula of the hormone molecule according to the IUPAC notion; (v)SMILE: give the sequence in SMILES format; (vi)Receptor: The corresponding receptor to which the hormone binds; (viii)PDB ID; (viii)PubChem ID; (ix)KEGG ID; (x)HMDB ID

Receptors - includes (i)'Organism' and (ii)'Taxonomy' of the source organism; (iii)Subcellular location: tells if the the hormone is secreted or the specific location of the hormone in the cell; (iv)Post translational modification: it gives the information if any other molecule has been attached to the mature peptide like phosphorylation; (v)Function: gives a brief function of the hormone; (vi)Hormone: The corresponding hormone to which thereceptor binds; (viii)Uniprot ID

ORGANISATION OF DATABASE

Browse - In Hmrbase 2.0, data can be browsed by (i) Peptide hormones – this option displays information about descriptive properties of peptide hormones in the fields including 'Taxonomy ' and 'Organism' of the source organism, 'Sequence' representing the amino acid sequence of the mature peptide and the corresponding 'Receptor' (ii) Non-Peptide hormones – this option displays fields corresponding to non-peptide hormones including other 'Synonyms' used for the hormone, its 'Formula', 'IUPAC' and 'SMILES' structure, and corresponding 'Receptor' (iii) Receptors – this option displays the field including the source 'Organism' and its 'Taxonomy', 'Subcellular Location' of the receptor, and the corresponding 'Hormone'.

Basic Search: There are three types of Basic Search pages available: one for peptide hormones, one for non-peptide hormones, and one for receptors. These pages have the same layout, but the available data fields differ. The search function allows users to search for specific keywords within any or all data fields in the database. Users can also specify which search results they want to see.

Advance Search: This tool enables users to perform advanced searches using a variety of fields and complex search criteria, including boolean operators like AND, OR, and NOT. Users can retrieve, copy, and download the search results in various formats, including CSV and Excel. There are separate pages for peptide hormones, non-peptide hormones, and receptors. The data for each hormone and hormone-receptor is organised into panels containing detailed cards linked to their identification number.

Similarity: This feature allows users to search for similar sequences using the Basic Local Alignment Search Tool (BLAST) and Smith-Waterman algorithms. To use these tools, the user submits a FASTA format peptide sequence using the default or specified settings, and the server will automatically run a BLAST search against stored data. The Smith-Waterman algorithm also searches for peptides based on similarity.

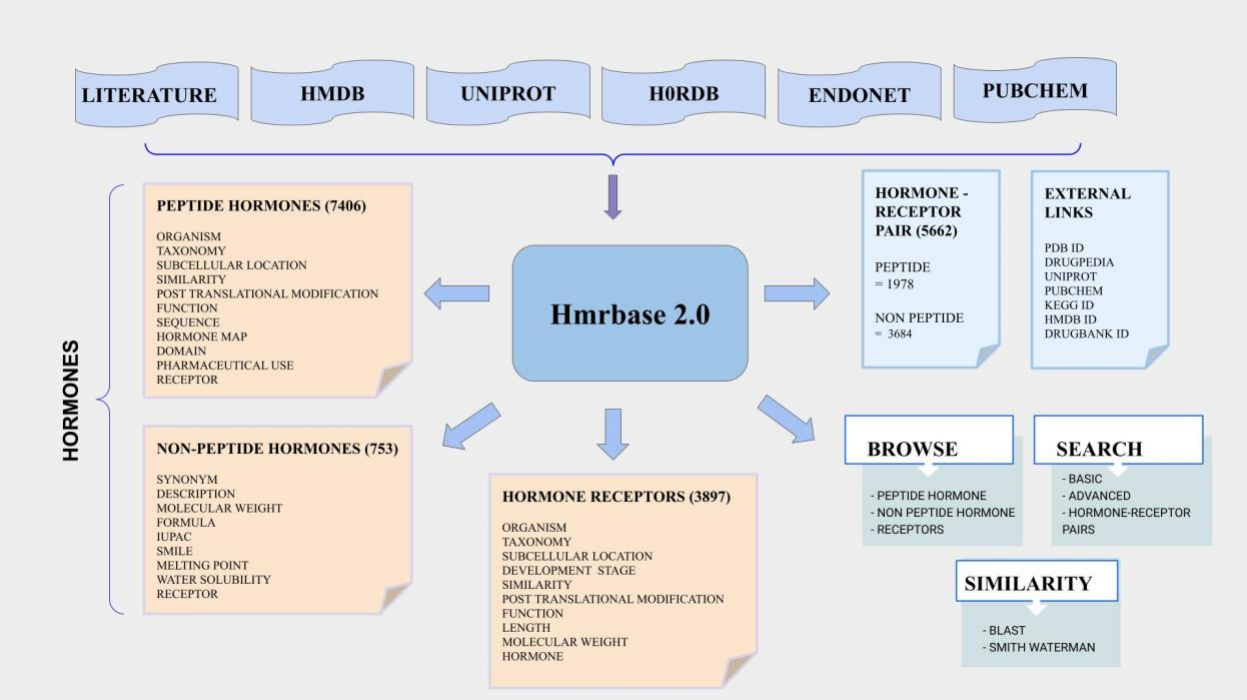

Figure 2: Architecture of Hmrbase 2.0

DATABASE STATISTICS

Hmrbase2, the upgraded edition of the Hmrbase database, contains 12056 total entries that were taken from publications, websites, and other databases. The 12056 entries include 3897 entries for hormone receptors, 7406 entries for peptide hormones, 753 entries for non-peptide hormones.

Hmrbase 2 is a database that provides details about hormones and their receptors from various organisms, including animals and plants. The database includes both peptide hormones, which are made up of small chains of amino acids, and non-peptide hormones. The database includes information about the amino acid sequences of mature peptide hormones and the SMILES (a chemical notation system) of non-peptide hormones. The database also includes links to other databases where users can find more information about the hormones and receptors, as well as structural information like PDB IDs and Pubchem IDs. The database also includes information about hormone receptor pairs, including 1978 for peptide hormones and 3684 for non-peptide hormones.

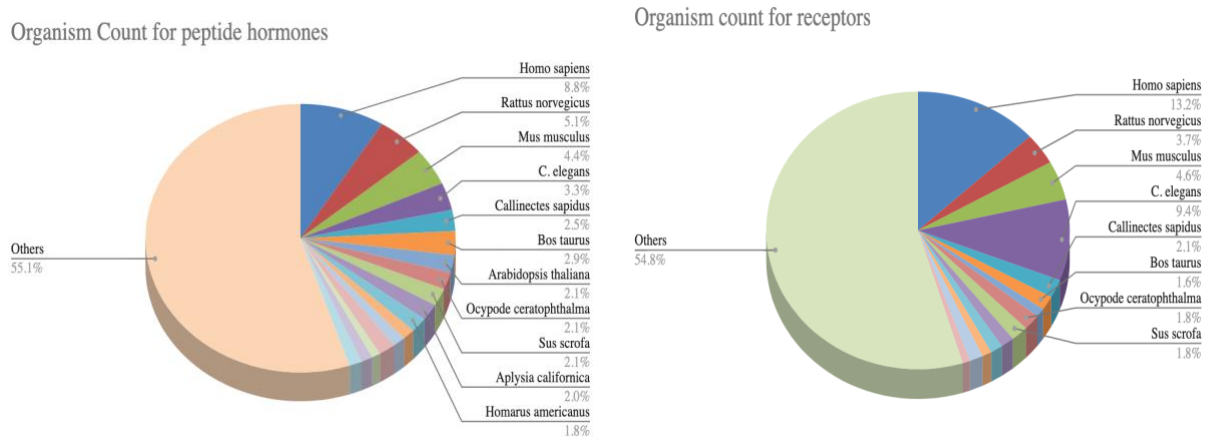

Figure 3: Distribution of source of peptide hormones (A) and hormone receptors (B).

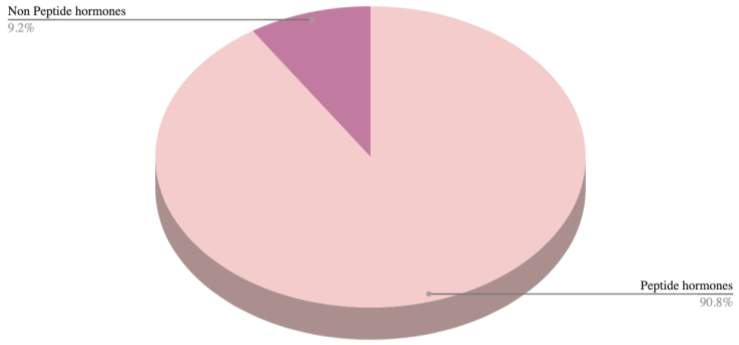

Figure 4: Proportion of peptide hormones and non-peptide hormones

COMPARISON WITH THE PREVIOUS VERSION

Hmrbase is a database that contains information about peptide hormones, non-peptide hormones, and their receptors. The original version of Hmrbase, released in 2009, contained three tables: one for peptide hormones, one for non-peptide hormones, and one for receptors. The peptide hormone table contained 1585 entries with information such as the receptor's name, source organism, sequence, length, and function. The non-peptide hormone table contained 370 entries with information such as name, function, IUPAC, SMILES, and receptor (29). The receptor table contained 2996 entries with information about the receptors, including their names, functions, lengths, and post-translational modifications. The updated version of Hmrbase, released later, contained more entries for hormones and their receptors and a total of 12056 entries. It also contained information about the pharmaceutical use of peptide hormones and links to Swiss-Prot, PubChem, and PDB for more details. The updated version also contained 5662 hormone receptor pairs compared to the 4121 in the original version.

Table 1 Hmrbase database comparison between two versions

	Peptide hormone		Non-peptide hormone		Total	
Data types	**Hmrbase**	**Hmrbase 2.0**	**Hmrbase**	**Hmrbase 2.0**	**Hmrbase**	**Hmrbase 2.0**
Hormone	1585	7406	370	752	1955	8158
Receptors	828	1597	2168	2300	2996	3897
Hormone-Receptor pairs	569	1978	3552	3684	4121	5662

COMPARISON WITH OTHER DATABASES

There are only a few databases available that provide information about hormones and their receptors. Hmrbase 2 is the only database that provides comprehensive information about peptide hormones, non-peptide hormones, and their receptors. HORDB is a database maintained by Dr. Zheng's team that only contains information about peptide hormones (30), while Endonet focuses on the intracellular network of molecules (31). Hmrbase 2 has incorporated data from Endonet and HORDB to create a comprehensive collection of all hormones and their receptors for use by the scientific community. The goal of Hmrbase 2 is to create a comprehensive resource for researchers working in this field.

Table 2. Comparison of Hmrbase 2.0 with the existing resources

Database	Peptide hormones	Non-peptide hormones	Hormone Receptors	Last updated
Hmrbase 2.0	**7406**	**752**	**3897**	**November 2022**
Hmrbase	1585	370	2996	2009
Endonet	447	29	615	2014
HorDB	5729	-	-	April 2022

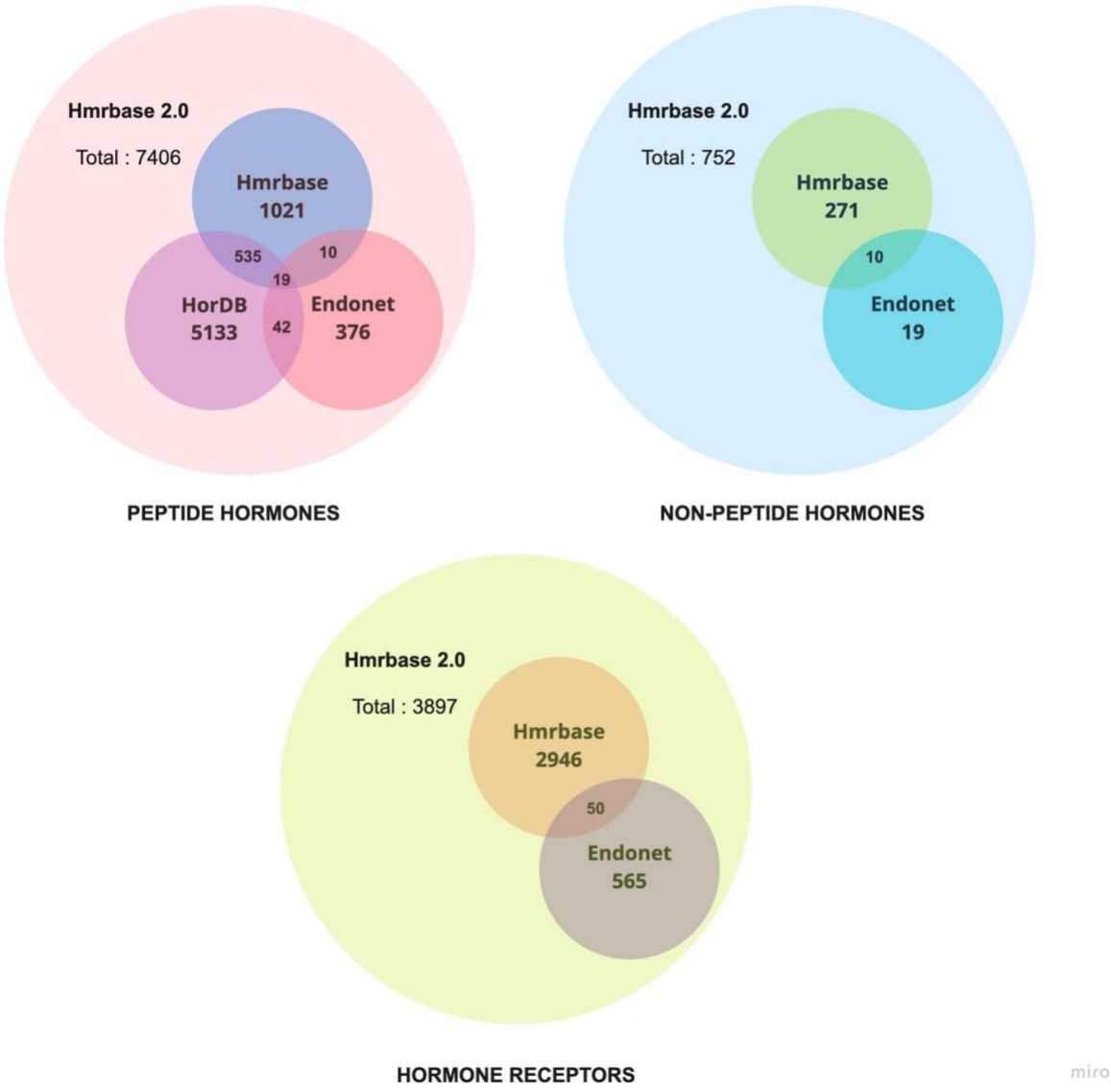

Figure 5: Proportion of information taken from various resources in Hmrbase 2

DISCUSSION

The Hmrbase 2 database is a comprehensive resource on hormones and their receptors that provides users with detailed information on the interactions between these molecules. It allows users to search for information on hormones and receptors using keywords and offers options for starting a search with either the hormone or the receptor. The database is user-friendly and provides detailed information on each hormone and receptor entry, including the precursor protein sequences for hormones. Both peptide and non-peptide hormones have been included in the database. The number of peptide

hormones is way more than that of non-peptide hormones. Peptides are typically smaller and more easily synthesised than steroid molecules. This allows them to be produced in larger quantities and distributed more widely throughout the body. Steroid molecules, on the other hand, are typically more extensive and more complex, and they require more specialised synthesis pathways. As a result, they are produced in smaller quantities and tend to be more concentrated in specific tissues or organs. It is helpful for both theoretical and practical endocrinologists and can be used to study the binding affinity of hormones to receptors or to develop new ligands. However, the database is not entirely error-free, and some hormone and receptor sequences may not be available. It will be updated approximately every three years.

Chapter - 3

Prediction of peptide hormones

INTRODUCTION

Peptide hormones are important regulatory molecules encoded in the genome and play a crucial role in transmitting specific information between cells and organs. These hormones are made from small chains of amino acids known as peptides. Peptide hormones are produced by a variety of organs and tissues in the body, including the pituitary gland, thyroid gland, and pancreas. This type of molecular communication has evolved over time into a sophisticated system for regulating growth, development, and homeostasis (32). Hormone imbalances are caused by various factors, including autoimmune disorders, genetic abnormalities, and damage to glands, and can lead to endocrine disorders such as diabetes and endocrine neoplasia (33). Therefore, it is crucial to focus on the therapeutics related to hormonal imbalances.

Peptide therapeutics are medications made from small chains of amino acids known as peptides (34). Peptides are commonly used as therapeutic agents because they can be easily synthesised and modified to target specific proteins or receptors in the body. Peptide therapeutics are used to treat a wide range of medical conditions, including cancer, diabetes, cardiovascular disease, and autoimmune disorders. There are several different types of peptide therapeutics, including peptide vaccines, which are used to stimulate the immune system to produce antibodies against specific pathogens or cancer cells (35-36). Peptide drugs are designed to target specific proteins or receptors in the body and are used to treat a wide range of medical conditions. Peptide hormones mimic the action of naturally occurring hormones in the body and are used to treat hormonal imbalances or deficiencies

Peptide hormones, like other peptides, bind to specific receptors on the surface of cells and trigger intracellular effects (37). They are attractive for drug development due to their specificity, target specificity, and other inherent qualities. Some peptide medications, such as insulin, can be used as replacement therapies to restore or supplement endogenous peptide hormone levels. This allows for therapeutic intervention miming natural pathways, as peptides are intrinsic signalling molecules for various physiological processes. Peptide hormones have been used as therapeutic agents for various conditions. Some examples of conditions for which peptide hormones have been used as therapeutics include

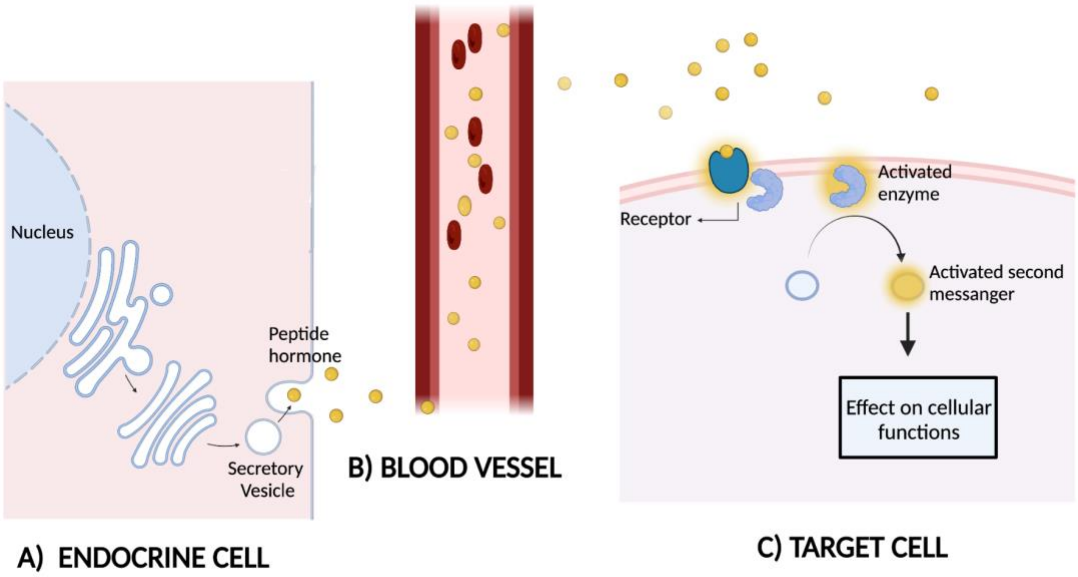

Figure 6: Mechanism of action of peptide hormones

Growth hormone deficiency: Growth hormone (GH) is a peptide hormone produced by the pituitary gland that stimulates growth and development. GH deficiency can occur due to various factors, including pituitary gland damage or defects in GH production. GH deficiency can lead to problems with growth and development, particularly in children and adolescents. GH therapy can be used to treat GH deficiency and promote average growth and development (38).

Diabetes: Insulin is a peptide hormone produced by the pancreas that regulates blood sugar levels by promoting the uptake and storage of glucose in cells. Insulin deficiency or resistance can lead to high blood sugar levels and the development of diabetes. Insulin therapy can be used to treat diabetes and control blood sugar levels (39).

Infertility: Follicle-stimulating hormone (FSH) and luteinising hormone (LH) are peptide hormones produced by the pituitary gland that regulates the development and function of the reproductive system. FSH and LH therapy can treat infertility caused by hormonal imbalances or deficiencies (40).

Osteoporosis: Parathyroid hormone (PTH) is a peptide hormone produced by the parathyroid gland that regulates the balance of calcium and phosphate in the body. PTH therapy can treat osteoporosis, a condition characterised by low bone density and an increased risk of fractures .

Peptide hormones are often used as therapeutics in the form of injectable drugs. They are usually administered by injection because they are broken down by digestive enzymes in the gut and are not absorbed when taken orally. The use of peptide hormones as therapeutics can be associated with side effects, and it is essential to monitor patients receiving peptide hormone therapy carefully.

For the prediction and creation of therapeutic peptides, a variety of techniques have been created, including TPpred-ATMV (41), THPep (42), AntiCP2.0 (43), and PrMFTP (44). However, there is currently no technique specifically designed to predict peptide hormones. In this study, we propose an approach called HOPPred to classify peptide sequences as either hormone or non-hormone.

MATERIALS AND METHODS

Dataset compilation and preprocessing

We sourced 5729 peptide hormone sequences from the Hmrbase2 database (45). These sequences are the mature form of the hormones, without the signal or precursor sequences. We removed any duplicate sequences, as well as those that were either too long or too short. The resulting positive dataset consists of 1174 peptide hormones from both plants and animals. To obtain a negative dataset, we used the peptide atlas repository to select 1174 non-hormone peptides that were similar in length to the positive dataset. The amino acid composition of the negative dataset was checked to ensure it was similar to the average amino acid composition of proteins found in nature. In the end, we had a positive dataset of 1174 peptide hormones and a negative dataset of 1174 non-hormone peptides, with zero per cent similarity between the two.

Feature generation

We generated over 9000 features using a method called Pfeature (46). We calculated 9529 features, which were used by models for the predictions. The details and length of each feature are listed in Table 3.

Table 3: List of all the computed features along with their vector length

Name of the Feature	Feature vector length
Amino acid composition (AAC)	20
Amphiphilic pseudo amino acid composition (APAAC)	23
Atom composition (ATC)	5
Bond composition (BTC)	4
Composition enhanced Transition Distribution (CTD)	189
Conjoint Triad Calculation (CTC)	343
Dipeptide composition (DPC)	400
Distance distribution of residue (DDOR)	20
Physicochemical Properties Composition (PCP)	30
Pseudo amino acid composition (PAAC)	21
Quasi-sequence order (QSO)	42
Residue Repeat Information (RRI)	20
Shannon Entropy of Physicochemical Property (SPC)	25
Shannon Entropy of Residues (SER)	20
Shannon-Entropy of Protein (SEP)	1
Tripeptide composition (TPC)	8000

Similarity search using BLAST

To identify peptide hormones, we used the BLAST shortp tool, which is a commonly used method for annotating protein and nucleotide sequences (47). We created a database of known hormone and non-hormone sequences and searched the query sequences against it. Based on the similarity of the peptide sequences to the sequences in the database, we identified peptide hormones using the top hit at various E-value cutoffs. This method has been well-studied and used in previous research (48-49).

Motif analysis using MERCI

The MERCI tool is used to identify recurring patterns, called motifs, in a set of peptide sequences (50). This tool utilises a Perl script and default settings to analyse the sequences and can provide useful information about peptide hormone sequences.

Feature selection

To identify the most important features from a large number of features, we used the Recursive Feature Elimination (RFE) method from the Scikit-learn package in Python, with Logistic Regression as the estimator (51). The features were chosen from standardised data using the StandardScaler from the Scikit-learn package (52). RFE removes the weakest features from the set until a certain number of features is reached. Unlike other feature selection methods focusing on individual feature properties, RFE focuses on features that impact model performance (53). We used the top 50 most relevant features in the machine-learning models.

Machine learning models

We used several machine-learning algorithms, such as decision tree, random forest, logistic regression, k-nearest neighbours, Gaussian Naive Bayes, and XGBoost, to classify peptide hormones and non-hormonal peptides using the Scikit-learn toolkit in Python. The decision tree algorithm learns decision rules from input to predict the output (54), the random forest classifier uses multiple decision trees to create a single prediction (55), logistic regression calculates the probability of an event using a logistic function (56), k-nearest neighbours predicts the class that is most commonly chosen by the closest data points (57), Gaussian Naive Bayes is a probabilistic classifier based on Bayes' theorem (58), and XGBoost makes predictions iteratively (59).

Deep Learning Techniques

1. Tabnet

   This text describes the use of a deep learning model called TabNet for tabular learning. TabNet uses sequential attention to choose which features to consider at each stage of the model. This method is advantageous because it allows for instance-wise feature selection, meaning that each row of data may have different features selected. This is in contrast to a model like XGBoost, which does not have this capability (60).

2. TextCNN

   In this study, we used a machine learning model called TextCNN to classify text data. To do this, we transformed the text into sequences of words with a length of 3 and used these sequences as input to the TextCNN model. The embedding dimension of the model was 8000, with 20 possible alphabets at each position in the sequence. We also padded the sequences to

ensure that the word lengths were consistent. The output of the TextCNN model was passed through a Random Forest (RF) model to make a final prediction.

Five-fold cross-validation

To ensure that our machine learning models were not biased or overfitted to the data, we employed a 5-fold cross-validation procedure. The data was divided into training and validation sets in a 80:20 ratio. The machine learning models were evaluated using 5-fold cross-validation on 80% of the training data, while being kept unaware of the remaining 20%. This process involves dividing the training data into 5 folds, using 4 folds for training and the remaining fold as a test set for internal validation. This is repeated 5 times, allowing each fold to be the test fold once.

Evaluation parameters

To assess the effectiveness of the machine learning models, we used standard evaluation metrics, including sensitivity (sen), specificity (spec), accuracy (acc), Matthews correlation coefficient (mcc), and area under the receiver operating characteristic curve. The model's ability to predict hormones is measured by sensitivity (Equation 1), while specificity measures the model's ability to correctly predict non-hormones (Equation 2). Accuracy represents the proportion of successful predictions for both hormones and non-hormones (Equation 3), while the Matthews correlation coefficient (MCC) indicates the relationship between the predicted and observed values (Equation 4). The area under the receiver operating characteristic curve (AUC) is a threshold-independent parameter that shows the model's ability to distinguish between the two classes, with sensitivity on one axis and 1-specificity on the other.

$$Sensitivity = \frac{T_P}{T_P + F_N} \quad (i)$$

$$Specificity = \frac{T_N}{T_N + F_P} \quad (ii)$$

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (iii)$$

$$MCC = \frac{(T_P \times T_N) - (F_P \times F_N)}{\sqrt{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}} \quad (iv)$$

Where, T_P, T_N, F_P and F_N stand for true positive, true negative, false positive and false negative, respectively.

Hybrid approach

In this study, a hybrid approach was used to improve the prediction of the model. This approach involves using three different techniques to calculate a score for each peptide sequence: a similarity-based approach using BLAST, a motif-based approach using MERCI, and a machine learning approach. The results from each of these techniques are combined to determine the overall score for the peptide sequence, which is then used to classify it as either a hormone or a non-hormone peptide. The positive predictions (hormone peptides) are given a weight of "+0.5," the negative predictions (non-hormone peptides) are given a weight of "-0.5," and no hits are given a weight of "0." This hybrid approach allows for more accurate classification of the peptide sequences.

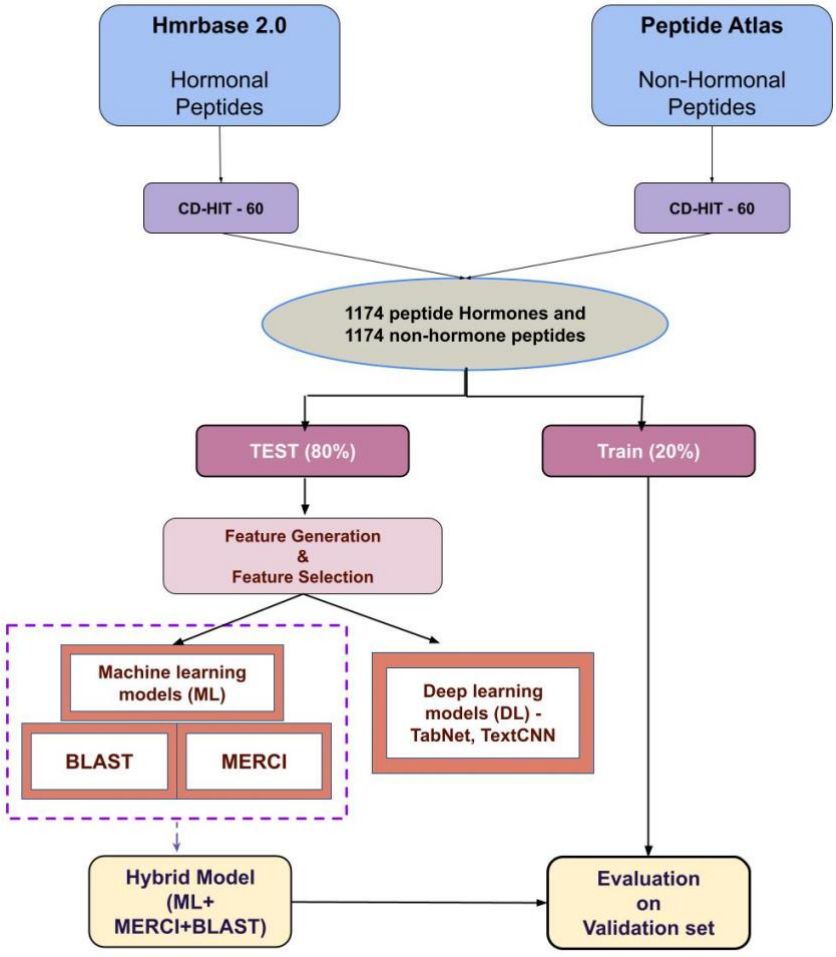

Figure 7: Schematic architecture of model for prediction of peptide hormones

RESULTS

Compositional analysis

The amino acid composition of peptide hormones was compared to that of random peptides and it was found that hormonal sequences tend to have higher levels of phenylalanine, glycine, proline, arginine, and serine compared to their average amounts found in nature, as shown in Figure 8.

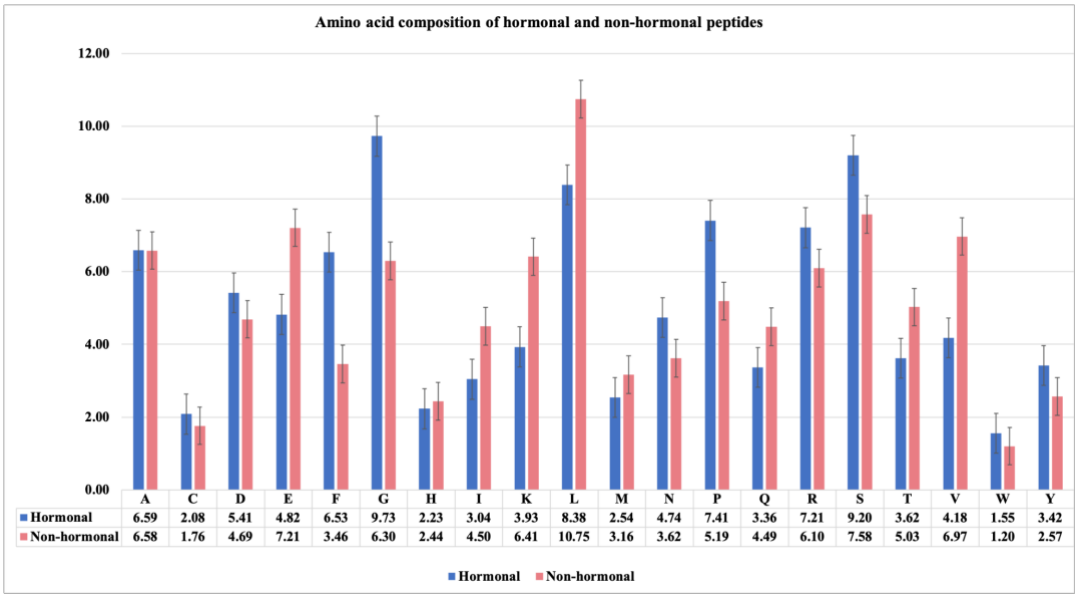

Figure 8: Amino acid composition analysis for peptide hormones and non-hormone peptides

Performance of prediction models

Composition-based features:

The random forest-based model was found to perform better than other models when using features such as amino acid composition to classify peptide hormones and non-hormonal peptide sequences. For the main dataset, it achieved an area under the receiver operating characteristic curve (AUC) of 0.88 and 0.892 for the training and validation datasets, respectively.

Table 4: The performance of ML-based models developed using amino acid composition

Amino Acid Composition (AAC)										
	Training					Validation				
Model	Sens(%)	Spec(%)	Acc(%)	AUC	MCC	Sens(%)	Spec(%)	Acc(%)	AUC	MCC
RF	80.786	80.342	80.564	0.888	0.611	79.31	84.034	81.702	0.892	0.634

LR	72.187	73.077	72.63	0.808	0.453	74.569	70.168	72.34	0.797	0.448
XGB	80.679	80.235	80.458	0.877	0.609	77.586	81.933	79.787	0.875	0.596
KNN	62.102	63.248	62.673	0.68	0.254	71.121	63.025	67.021	0.719	0.342
GNB	71.231	70.62	70.927	0.787	0.419	75.431	71.429	73.404	0.787	0.469
DT	66.667	67.201	66.933	0.72	0.339	60.776	66.387	63.617	0.698	0.272
SVC	74.098	72.329	73.216	0.807	0.464	76.724	68.487	72.553	0.792	0.453

ML Models using selected features:

The LR model performed the best out of all the models tested on both the main and alternate datasets when using a reduced number of features. The performance of these models is shown in Table 5

Table 5: The performance of ML-based models developed using selected features

RFE Top30										
	Training					**Validation**				
Model	**Sens(%)**	**Spec(%)**	**Acc(%)**	**AUC**	**MCC**	**Sens(%)**	**Spec(%)**	**Acc(%)**	**AUC**	**MCC**
RF	80.679	80.769	80.724	0.9	0.614	77.586	83.613	80.638	0.883	0.613
LR	**85.987**	**86.111**	**86.049**	**0.924**	**0.721**	**84.052**	**83.613**	**83.83**	**0.899**	**0.677**
XGB	81.104	81.517	81.31	0.898	0.626	80.172	80.252	80.213	0.89	0.604
KNN	77.813	76.603	77.21	0.859	0.544	76.724	76.471	76.596	0.836	0.532
GNB	65.393	66.453	65.921	0.738	0.318	78.448	56.303	67.234	0.755	0.356
RFE Top50										
	Training					**Validation**				
Model	**Sens(%)**	**Spec(%)**	**Acc(%)**	**AUC**	**MCC**	**Sens(%)**	**Spec(%)**	**Acc(%)**	**AUC**	**MCC**
RF	84.289	84.081	84.185	0.914	0.684	80.172	84.034	82.128	0.901	0.643
LR	**87.049**	**87.393**	**87.22**	**0.941**	**0.744**	**85.345**	**86.555**	**85.957**	**0.926**	**0.719**
XGB	84.183	84.295	84.239	0.918	0.685	77.586	84.454	81.064	0.898	0.622
KNN	56.369	46.795	51.597	0.518	0.032	57.328	47.899	52.553	0.565	0.052
GNB	24.522	95.94	60.117	0.758	0.292	22.845	97.059	60.426	0.829	0.298
DT	72.718	72.115	72.417	0.776	0.448	72.414	73.95	73.191	0.789	0.464

Deep Learning Techniques:

TextCNN

TextCNN is a deep learning method to perform text classification tasks. We divided the sequences into the words of length = 3 and used them to build our deep learning model. This model was able to achieve an AUROC of 0.98 on the training dataset, however, it was not able to perform well on the testing dataset giving an AUROC of 0.90. The MCC was also seen to decrease from 0.88 in training dataset to 0.67 in testing dataset.

TabNet

TabNet is a deep learning method that uses tabular data as input and is trained using gradient-descent based optimization. The 9149 features generated from the Pfeature were passed as features through the TabNet model to get the train AUROC as 0.81 and the validation AUROC as 0.75 with an MCC of 0.61 for training dataset and 0.57 for validation dataset. The performances of the deep learning models – TextCNN and TabNet have been summarised in Table 6.

Table 6: The performance of and deep learning models.

Deep learning methods										
Model	**Training**					**Validation**				
	Sens(%)	**Spec(%)**	**Acc(%)**	**AUC**	**MCC**	**Sens(%)**	**Spec(%)**	**Acc(%)**	**AUC**	**MCC**
Text CNN	96	93	94	0.98	0.88	87	79	83	0.90	0.67
Tabnet	79	80	80	0.81	0.61	73	75	74	0.75	0.57

Motif-based approach

The combination of the motif-based approach from MERCI with the composition-based models developed using different machine learning techniques was evaluated for both datasets. The results are shown in Table 7.

Table 7: The performance of ML-based models in combination with MERCI

ML + MERCI										
Model	**Training**					**Validation**				
	Sens(%)	**Spec(%)**	**Acc(%)**	**AUC**	**MCC**	**Sens(%)**	**Spec(%)**	**Acc(%)**	**AUC**	**MCC**

RF	84.5	83.55	84.03	0.92	0.68	81.03	86.13	83.62	0.92	0.67
LR	**87.9**	**87.93**	**87.91**	**0.95**	**0.76**	**86.64**	**87.39**	**87.02**	**0.94**	**0.74**
XGB	85.03	84.51	84.77	0.92	0.7	80.17	85.29	82.77	0.91	0.66
KNN	58.39	52.46	55.43	0.59	0.11	58.19	57.56	57.87	0.67	0.16
GNB	26.86	96.05	61.34	0.78	0.32	25.43	97.9	62.13	0.85	0.34
DT	73.78	73.82	73.8	0.8	0.48	64.22	75.21	69.79	0.76	0.4

BLAST performance

In this study, we performed a BLAST search to identify hormonal sequences based on similarity between the two sequences. We used the standard top hit approach on various e-values.

Table 8. BLAST-based search on training and validation dataset (here, Chits = correct hits, Whits = wrong hits, Nhits = no hits)

E-value	Training						Validation					
	Hormonal			Non-hormonal			Hormonal			Non-hormonal		
	Chits	Whits	Nhits	Chits	Whits	Nhits	Chits	Whits	Nhits	Chits	Whits	Nhits
10^{-6}	269 (28.56%)	0 (0.00%)	673 (71.44%)	0 (0.00%)	0 (0.00%)	936 (100%)	65 (28.02%)	0 (0.00%)	167 (71.98%)	0 (0.00%)	0 (0.00%)	238 (100%)
10^{-5}	302 (32.06%)	0 (0.00%)	640 (67.94%)	4 (0.43%)	0 (0.00%)	932 (99.57%)	78 (33.62%)	0 (0.00%)	154 (66.38%)	0 (0.00%)	0 (0.00%)	238 (100%)
10^{-4}	350 (37.15%)	0 (0.00%)	592 (62.85%)	4 (0.43%)	0 (0.00%)	932 (99.57%)	90 (38.79%)	0 (0.00%)	142 (61.21%)	0 (0.00%)	0 (0.00%)	238 (100%)
10^{-3}	427 (45.33%)	0 (0.00%)	515 (54.67%)	8 (0.85%)	0 (0.00%)	928 (99.15%)	107 (46.12%)	0 (0.00%)	125 (53.88%)	0 (0.00%)	0 (0.00%)	238 (100%)
10^{-2}	527 (55.94%)	0 (0.00%)	415 (44.06%)	8 (0.85%)	0 (0.00%)	928 (99.15%)	126 (54.31%)	0 (0.00%)	106 (45.69%)	2 (0.84%)	0 (0.00%)	236 (99.16%)

10^{-1}	628 (66.67%)	1 (0.11%)	313 (33.23%)	8 (0.85%)	1 (0.11%)	927 (99.04%)	144 (62.07%)	0 (0.00%)	88 (37.93%)	2 (0.84%)	0 (0.00%)	236 (99.16%)
1	705 (74.84%)	3 (0.32%)	234 (24.84%)	21 (2.24%)	10 (1.07%)	905 (96.69%)	163 (70.26%)	4 (1.72%)	65 (28.02%)	10 (4.2%)	2 (0.84%)	226 (94.96%)
10^{1}	766 (81.32%)	40 (4.25%)	136 (14.44%)	165 (17.63%)	114 (12.18%)	657 (70.19%)	180 (77.59%)	19 (8.19%)	33 (14.22%)	51 (21.43%)	34 (14.29%)	153 (64.29%)
10^{2}	754 (80.04%)	179 (19.00%)	9 (0.96%)	520 (55.56%)	374 (39.96%)	42 (4.49%)	176 (75.86%)	55 (23.71%)	1 (0.43%)	140 (58.82%)	89 (37.39%)	9 (3.78%)

Hybrid approach

To improve the accuracy of predicting virulent proteins, we used a combination of different approaches. These approaches included a BLAST search based on similarity, a motif approach using MERCI, and a machine learning model based on amino acid composition (AAC). We first used ensemble BLAST and MERCI to classify protein sequences, and then used any remaining sequences to predict with the AAC-based machine learning model. The performance of this hybrid approach was significantly better than using any individual method alone. The random forest machine learning model performed best for both datasets on the training and validation data, with an area under the curve (AUC) of 0.96 and 0.97 for the main dataset and 0.92 and 0.94 for the alternate dataset.

Table 9: The performance of ML-based models in combination with MERCI and BLAST

ML + Motif-search + BLAST										
	Training					**Validation**				
Model	**Sens (%)**	**Spec (%)**	**Acc (%)**	**AUC**	**MCC**	**Sens (%)**	**Spec (%)**	**Acc (%)**	**AUC**	**MCC**
RF	88.96	85.90	87.43	0.95	0.75	86.21	88.66	87.45	0.95	0.75
LR	**91.19**	**90.38**	**90.79**	**0.97**	**0.82**	**90.09**	**89.5**	**89.79**	**0.96**	**0.80**
XGB	88.96	87.71	88.34	0.96	0.77	85.34	87.82	86.6	0.95	0.73
KNN	77.18	72.54	73.87	0.85	0.50	74.14	75.63	74.89	0.86	0.50
GNB	69.43	96.26	82.80	0.90	0.68	64.66	97.9	81.49	0.91	0.67
DT	81.95	79.70	80.83	0.89	0.62	77.16	76.89	77.02	0.86	0.54

Webserver and Standalone Software

The webserver for HOPPred (accessed at https://webs.iiitd.edu.in/raghava/hoppred/) that we created can distinguish between peptide hormones and non-hormone peptides. The front end and back end of the web server were created using HTML5, Java, CSS3, and PHP scripts. All of the most recent gadgets, including smartphones, tablets, iMacs, and desktop PCs, are all compatible with this web server. Predict, design, and peptide design modules are the primary modules on HOPPred.

DISCUSSION

Peptide hormones are a group of regulatory molecules that are encoded in the genome and play a crucial role in transferring information between cells and organs. Hormonal imbalances can lead to endocrine disorders, and it is important to focus on therapies for these disorders. Peptide hormones have an appealing pharmacological profile and can be specifically targeted to cells, making them a good starting point for developing new treatments. Some peptide medications act as replacement therapies, restoring or replacing peptide hormones when levels are insufficient or absent. In recent years, there has been a shift from replicating natural peptides to rationally designing peptides with desirable biochemical and physiological properties. There are several methods that have been proposed for predicting and designing therapeutic peptides, but none specifically for predicting peptide hormones. In this study, the authors developed a method called HOPPred to classify peptide hormone and non-hormone peptide sequences using machine learning algorithms. They also used a hybrid approach, which combines three techniques to calculate a score: a similarity-based approach using BLAST, a motif-based approach using MERCI, and a machine learning approach. The authors found that the hybrid approach improved the prediction performance and created a web server, HopPred, to make their method accessible to the scientific community. They hope that their prediction technique will be used to develop more accurate and efficient peptide-based treatments for various diseases.

Chapter - 4

Conclusion

Hormones are chemical messengers that play a crucial role in regulating various physiological processes in the body. Imbalances in hormones can lead to the development of various health conditions, such as cancer, osteoporosis, and diabetes. It is important to collect and organise information about hormones and their receptors. Hmrbase2.0 is a database that provides information about hormones and their receptors, including information about their type, source organism, sequence, and other chemical properties. It is designed for endocrinologists and includes information about precursor protein sequences and protein domains, which can be used to study binding affinity and design better ligands for specific receptors. Hmrbase2.0 is easy to use and provides a comprehensive view of hormones and their receptors that is not currently available on any other website.

Peptide therapeutics, including peptide vaccines, drugs, and hormones, are used to treat various medical conditions. Peptide drugs, which are specific and selective compared to traditional small molecule therapeutics, can be improved by using computational approaches to understand their mechanisms. Several methods exist for predicting and designing therapeutic peptides, but none specifically for predicting peptide hormones. In this study, a technique called HOPPred was developed to classify peptide sequences as hormone or non-hormone.

Altogether, the research presented in this thesis provides some unique approaches for properly annotating essential molecules of the endocrine system. We anticipate that clinicians and researchers will use the findings of our investigations to develop advanced cancer treatment approaches.

Bibliography

1. Stárka, L., & Dušková, M. (2020). What is a hormone?. Physiological research, 69(Suppl 2), S183–S185.
2. Kraemer, W. J., Ratamess, N. A., Hymer, W. C., Nindl, B. C., & Fragala, M. S. (2020). Growth Hormone(s), Testosterone, Insulin-Like Growth Factors, and Cortisol: Roles and Integration for Cellular Development and Growth With Exercise. Frontiers in endocrinology, 11, 33.
3. Wallen K. (2020). Hormones and Behavior: An emergent journal. Hormones and behavior, 121, 104692.
4. Loumaye, E., Thorner, J., & Catt, K. J. (1982). Yeast mating pheromone activates mammalian gonadotrophs: evolutionary conservation of a reproductive hormone?. Science (New York, N.Y.), 218(4579), 1323–1325.
5. Leblebicioglu, B., Connors, J., & Mariotti, A. (2013). Principles of endocrinology. Periodontology 2000, 61(1), 54–68.
6. Alexander, S. P. H., Christopoulos, A., Davenport, A. P., Kelly, E., Mathie, A., Peters, J. A., Veale, E. L., Armstrong, J. F., Faccenda, E., Harding, S. D., Pawson, A. J., Sharman, J. L., Southan, C., Davies, J. A., & CGTP Collaborators (2019). THE CONCISE GUIDE TO PHARMACOLOGY 2019/20: G protein-coupled receptors. British journal of pharmacology, 176 Suppl 1(Suppl 1), S21–S141.
7. Wright, J. R., Jr, Yang, H., Hyrtsenko, O., Xu, B. Y., Yu, W., & Pohajdak, B. (2014). A review of piscine islet xenotransplantation using wild-type tilapia donors and the production of transgenic tilapia expressing a "humanized" tilapia insulin. Xenotransplantation, 21(6), 485–495.
8. Tokarz, V. L., MacDonald, P. E., & Klip, A. (2018). The cell biology of systemic insulin function. The Journal of cell biology, 217(7), 2273–2289.
9. Norton, L., Shannon, C., Gastaldelli, A., & DeFronzo, R. A. (2022). Insulin: The master regulator of glucose metabolism. Metabolism: clinical and experimental, 129, 155142.
10. Ward, C. W., & Lawrence, M. C. (2009). Ligand-induced activation of the insulin receptor: a multi-step process involving structural changes in both the ligand and the receptor. BioEssays : news and reviews in molecular, cellular and developmental biology, 31(4), 422–434.
11. Gutmann, T., Kim, K. H., Grzybek, M., Walz, T., & Coskun, Ü. (2018). Visualization of ligand-induced transmembrane signaling in the full-length human insulin receptor. The Journal of cell biology, 217(5), 1643–1649.
12. Dimitriadis, G., Mitrou, P., Lambadiari, V., Maratou, E., & Raptis, S. A. (2011). Insulin effects in muscle and adipose tissue. Diabetes research and clinical practice, 93 Suppl 1, S52–S59

13. Bouskila, M., Hirshman, M. F., Jensen, J., Goodyear, L. J., & Sakamoto, K. (2008). Insulin promotes glycogen synthesis in the absence of GSK3 phosphorylation in skeletal muscle. American journal of physiology. Endocrinology and metabolism, 294(1), E28–E35.
14. Gustavsson, J., Parpal, S., & Strålfors, P. (1996). Insulin-stimulated glucose uptake involves the transition of glucose transporters to a caveolae-rich fraction within the plasma membrane: implications for type II diabetes. Molecular medicine (Cambridge, Mass.), 2(3), 367–372.
15. Guthrie, R. A., & Guthrie, D. W. (2004). Pathophysiology of diabetes mellitus. Critical care nursing quarterly, 27(2), 113–125.
16. Peramunage, D., & Nikravan, S. (2020). Anesthesia for Endocrine Emergencies. Anesthesiology clinics, 38(1), 149–163.
17. Miller, W. L., & Auchus, R. J. (2011). The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. Endocrine reviews, 32(1), 81–151.
18. Adams J. S. (2005). "Bound" to work: the free hormone hypothesis revisited. Cell, 122(5), 647–649.
19. Hammes, A., Andreassen, T. K., Spoelgen, R., Raila, J., Hubner, N., Schulz, H., Metzger, J., Schweigert, F. J., Luppa, P. B., Nykjaer, A., & Willnow, T. E. (2005). Role of endocytosis in cellular uptake of sex steroids. Cell, 122(5), 751–762.
20. Holst, J. P., Soldin, O. P., Guo, T., & Soldin, S. J. (2004). Steroid hormones: relevance and measurement in the clinical laboratory. Clinics in laboratory medicine, 24(1), 105–118.
21. Leyser H. M. (1998). Plant hormones. Current biology : CB, 8(1), R5–R7.
22. Asami, T., Nakano, T., & Fujioka, S. (2005). Plant brassinosteroid hormones. Vitamins and hormones, 72, 479–504.
23. Zhabinskii, V. N., Khripach, N. B., & Khripach, V. A. (2015). Steroid plant hormones: effects outside plant kingdom. Steroids, 97, 87–97.
24. Kooistra, A. J., Mordalski, S., Pándy-Szekeres, G., Esguerra, M., Mamyrbekov, A., Munk, C., Keserű, G. M., & Gloriam, D. E. (2021). GPCRdb in 2021: integrating GPCR sequence, structure and function. Nucleic acids research, 49(D1), D335–D343.
25. Dagan-Wiener, A., Di Pizio, A., Nissim, I., Bahia, M. S., Dubovski, N., Margulis, E., & Niv, M. Y. (2019). BitterDB: taste ligands and receptors database in 2019. Nucleic acids research, 47(D1), D1179–D1185.
26. Réau, M., Lagarde, N., Zagury, J. F., & Montes, M. (2019). Nuclear Receptors Database Including Negative Data (NR-DBIND): A Database Dedicated to Nuclear Receptors Binding

Data Including Negative Data and Pharmacological Profile. Journal of medicinal chemistry, 62(6), 2894–2904.

27. Singh, S., Chaudhary, K., Dhanda, S. K., Bhalla, S., Usmani, S. S., Gautam, A., Tuknait, A., Agrawal, P., Mathur, D., & Raghava, G. P. (2016). SATPdb: a database of structurally annotated therapeutic peptides. Nucleic acids research, 44(D1), D1119–D1126.
28. Panyayai, T., Ngamphiw, C., Tongsima, S., Mhuantong, W., Limsripraphan, W., Choowongkomon, K., & Sawatdichaikul, O. (2019). FeptideDB: A web application for new bioactive peptides from food protein. Heliyon, 5(7), e02076.
29. Rashid, M., Singla, D., Sharma, A., Kumar, M., & Raghava, G. P. (2009). Hmrbase: a database of hormones and their receptors. BMC genomics, 10, 307
30. Zhu, N., Dong, F., Shi, G., Lao, X., & Zheng, H. (2022). HORDB a comprehensive database of peptide hormones. Scientific data, 9(1), 187.
31. Dönitz, J., & Wingender, E. (2014). EndoNet: an information resource about the intercellular signaling network. BMC systems biology, 8, 49.
32. Sandoval, D. A., & D'Alessio, D. A. (2015). Physiology of proglucagon peptides: role of glucagon and GLP-1 in health and disease. Physiological reviews, 95(2), 513–548.
33. Thakur S. S. (2021). Proteomics and its application in endocrine disorders. Biochimica et biophysica acta. Proteins and proteomics, 1869(11), 140701.
34. Erak, M., Bellmann-Sickert, K., Els-Heindl, S., & Beck-Sickinger, A. G. (2018). Peptide chemistry toolbox - Transforming natural peptides into peptide therapeutics. Bioorganic & medicinal chemistry, 26(10), 2759–2765.
35. Drucker D. J. (2020). Advances in oral peptide therapeutics. Nature reviews. Drug discovery, 19(4), 277–289.
36. Timur, S. S., & Gürsoy, R. N. (2021). The role of peptide-based therapeutics in oncotherapy. Journal of drug targeting, 29(10), 1048–1062.
37. Lau, J. L., & Dunn, M. K. (2018). Therapeutic peptides: Historical perspectives, current development trends, and future directions. Bioorganic & medicinal chemistry, 26(10), 2700–2707.
38. Kato, Y., Murakami, Y., Sohmiya, M., & Nishiki, M. (2002). Regulation of human growth hormone secretion and its disorders. Internal medicine (Tokyo, Japan), 41(1), 7–13.
39. Wang, L., Wang, N., Zhang, W., Cheng, X., Yan, Z., Shao, G., Wang, X., Wang, R., & Fu, C. (2022). Therapeutic peptides: current applications and future directions. Signal transduction and targeted therapy, 7(1), 48.

40. Santi, D., Crépieux, P., Reiter, E., Spaggiari, G., Brigante, G., Casarini, L., Rochira, V., & Simoni, M. (2020). Follicle-stimulating Hormone (FSH) Action on Spermatogenesis: A Focus on Physiological and Therapeutic Roles. Journal of clinical medicine, 9(4), 1014.
41. 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. Bioinformatics (Oxford, England), 38(10), 2712–2718.
42. 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. Bioinformatics (Oxford, England), 38(10), 2712–2718.
43. Agrawal, P., Bhagat, D., Mahalwal, M., Sharma, N., & Raghava, G. P. S. (2021). AntiCP 2.0: an updated model for predicting anticancer peptides. Briefings in bioinformatics, 22(3), bbaa153.
44. Yan, W., Tang, W., Wang, L., Bin, Y., & Xia, J. (2022). PrMFTP: Multi-functional therapeutic peptides prediction based on multi-head self-attention mechanism and class weight optimization. PLoS computational biology, 18(9), e1010511.
45. Kaur, D., Arora, A., Patiyal, S., & Raghava, G. P. (2023). Hmrbase2: A comprehensive database of hormones and their receptors. bioRxiv 2023.01.31.526433.
46. Pande, A., Patiyal, S., Lathwal, A., Arora, C., Kaur, D., Dhall, A., Mishra, G., Kaur, H., Sharma, N., Jain, S., Usmani, S. S., Agrawal, P., Kumar, R., Kumar, V., & Raghava, G. P. S. (2023). Pfeature: A Tool for Computing Wide Range of Protein Features and Building Prediction Models. Journal of computational biology : a journal of computational molecular cell biology, 30(2), 204–222.
47. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403–410.
48. Saha, S., & Raghava, G. P. (2006). AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic acids research, 34(Web Server issue), W202–W209.
49. Sharma, N., Patiyal, S., Dhall, A., Pande, A., Arora, C., & Raghava, G. P. S. (2021). AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes. Briefings in bioinformatics, 22(4), bbaa294.
50. Vens, C., Rosso, M. N., & Danchin, E. G. (2011). Identifying discriminative classification-based motifs in biological sequences. Bioinformatics (Oxford, England), 27(9), 1231–1238.
51. Ren, K., Zeng, Y., Cao, Z., & Zhang, Y. (2022). ID-RDRL: a deep reinforcement learning-based feature selection intrusion detection model. Scientific reports, 12(1), 15370.

52. Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. Frontiers in neuroinformatics, 8, 14.
53. Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. Frontiers in neuroinformatics, 8, 14.
54. Flayer, C. H., Perner, C., & Sokol, C. L. (2021). A decision tree model for neuroimmune guidance of allergic immunity. Immunology and cell biology, 99(9), 936–948.
55. Yi, Y., Sun, D., Li, P., Kim, T. K., Xu, T., & Pei, Y. (2022). Unsupervised random forest for affinity estimation. Computational visual media, 8(2), 257–272.
56. Stoltzfus J. C. (2011). Logistic regression: a brief primer. Academic emergency medicine : official journal of the Society for Academic Emergency Medicine, 18(10), 1099–1104.
57. Miao, Y., Hunter, A., & Georgilas, I. (2021). An Occupancy Mapping Method Based on K-Nearest Neighbours. Sensors (Basel, Switzerland), 22(1), 139.
58. Joshi, D., Mishra, A., & Anand, S. (2012). A naïve Gaussian Bayes classifier for detection of mental activity in gait signature. Computer methods in biomechanics and biomedical engineering, 15(4), 411–416.
59. Hou, N., Li, M., He, L., Xie, B., Wang, L., Zhang, R., Yu, Y., Sun, X., Pan, Z., & Wang, K. (2020). Predicting 30-days mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. Journal of translational medicine, 18(1), 462.
60. Arık, S. O. & Pfister, T. (2020). TabNet: Attentive Interpretable Tabular Learning. arXiv:1908.07442v5