Original Article

Comparative In-Silico Screening Of Potent Peptide Leads Using Docking Strategy & AI Approaches For The Treatment Of Liver Cancer

Abhishek[1], Avantika Goyal[1], Anjali Saini[1], Kandasamy Nagarajan[1*], Pankaj Bhatt[1*], Garima Kapoor[1], Sheena Mehta[1], Siddheshwari Mishra[2], Suyash Shrivastava[2], Anil Ahlawat[2], Ajay Shrivastava[2], Surbhi Kamboj[1]

[1]KIET School of Pharmacy, KIET Group of Institutions, 13 km milestone, Ghaziabad, Delhi-NCR, 201206, India
[2]KIET Group of Institutions, 13 km milestone, Ghaziabad, Delhi-NCR, 201206, India
Correspondence author email id: rknagaraaja@gmail.com
DOI: 10.47750/pnr.2023.14.03.394

Abstract

Hepatocellular carcinoma (HCC) or Liver cancer is a common type of tumor mainly found in patients that are diagnosed with cirrhosis. Our recent therapies for liver cancer include laser treatment, surgery, or standard medicines like sorafenib, Lenvatinib has major side effects like bleeding gums, blurred vision etc. so to prevent these types of side effect and make a more bio-familiar drug with body environment tried to make the peptide-based drug. Casein kinase 2 and Toll-like receptor 2 are selected as the target of our study, which are essential for cell growth and proliferation and act as a potent suppressor of apoptosis, offering an important link to their involvement in cancer, as deregulation of both cell proliferation and apoptosis belong among the key features of cancer cell biology. Various vital amino acids, of smaller-chain dipeptides were selected. The Swiss ADME tool was used to predict the pharmacokinetic properties of dipeptides such as GIT absorption, Lipinski, etc. and then they were subjected to Pro-tox II software to predict the toxicity. The Swiss online tool DOCK & CB dock-2 (PDB code: 4GRB) was used for docking, along with Chimera visualizer and Discovery studio for a detailed investigation of 3D interactions. As per the docking result, the Phe-Tyr was found to the be most promising dipeptide lead as a potent inhibitor of CK2 with a ΔG value of -9.06Kcal/mol followed by Phe-His value of ΔG = -8.44 in comparison to standard drug Lenvatinib and Sorafenib whose ΔG value are -8.40 & -9.37 respectively. In SVM results were some differences such as mostly dipeptides that have similar to standard drugs belonging to the Tryptophan family followed by proline and Phenylalanine. In the Gaussian classifier, the good-scoring peptides were Pro-Trp, Trp-Thr, Trp-Ser, and Trp-Ala. When comes to cosine similarity Pro-Ser followed by Pro-Thr, Pro-Leu & Pro-His considered the most promising lead dipeptides. Comparative studies found that there were some differences between man-made results & M.L results.

Keywords: Casein Kinase 2, Dipeptides, Docking, Lenvatinib, Liver cancer, Sorafenib, Support Vector Machine.

Introduction

Artificial intelligence is a field of computer science or IT that works on the basis of programming and using an input data file gives an output without distortion of a specific thing that gives an advantage over human thinking. Machine learning is the main part of artificial intelligence Machine learning is further divided into two branches: Supervised ML (Machines are trained using marked training data and based on this data machine provide output) and Unsupervised ML. Under Supervised ML, some methods such as Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) are also used in the current drug design scenario.

Peptides are a combination of amino acid chains that are arranged to form a polymeric structure and based on the number of amino acids that combine to form a peptide, they are called dipeptide, tripeptide, tetra peptide, etc. And generally, polypeptides are referred to as proteins. Amino acids consist of an amino group (-NH2) and a carboxyl group. Therapeutic peptides contribute significantly to the treatment of oncological, diabetic, cardiovascular, immunosuppressive and antifungal indications.

Chronic hepatitis is a major cause of hepatocellular carcinoma and this study was conducted to investigate tumor-

specific dipeptides using liquid chromatography tandem mass spectrometry [1]. Two-year ultrasound surveillance is recommended in these patients, as this diagnosis is possible at an early stage. Resection (the process of cutting out tissue) is best for a candidate with a solitary tumor and preserved liver function. Image-guided ablation is often used, but its effectiveness is limited by tumor size and location. Some standard drugs used for treatment are Sorafenib, Lenvatinib and Regorafenib. [2]. Histone deacetylase 2 enzyme plays an important role in the development of HCC through the regulation of cell cycle components and transcriptional levels and is related to CK2 overexpression. Thus, targeted or selective disruption of CK2 affected the cell cycle and arrested several liver cancer cells in the G2/M phase and suppressed growth through suppression of cyclin B and cdc25c in liver cancer cells [3]. TLR2 is expressed by various liver cells, including Kupffer cells (KCs), hepatocytes, and hepatic stellate cells (HSCs). Pathologically, the crosstalk between antigens and TLR2 may preferentially trigger a characteristic set of signaling mechanisms in these liver cells, thereby inducing the production of inflammatory and fibrogenic cytokines that can initiate and prolong liver inflammation, ultimately leading to fibrosis. [4]. Anticancer peptides (ACPs) are a series of short peptides of 10-60 amino acids that can inhibit the proliferation or migration of tumor cells or suppress the formation of tumor blood vessels and are less likely to cause drug resistance. However, ACPs can be degraded by proteases or in many cases lead to cytotoxicity. To overcome these shortcomings, much research has focused on reconstructing or modifying ACPs to improve their anticancer activity while reducing their cytotoxicity. ACP modification mainly includes main chain reconstruction and side chain modification [5]. Complex and big data from genomics, proteomics, microarray data and clinical trials are also a bottleneck in the drug discovery process, so time-consuming processes and more animal testing can be avoided using machine learning [6]. Emerging computational approaches, including computational biology, computer-aided drug design, and artificial intelligence, have the potential to accelerate the efficiency of drug discovery by minimizing time and financial costs [7]. We calculated available topological descriptors such as Lipinski's rule, Swiss ADME, Molinspiration and Pro-tox II as selective empirical drug filters to correlate their bio potency. The binding of the test set analog was calculated directly from the Swiss dock against various receptors. For screening, a free database of all molecules, such as 45 dipeptide combinations, should be developed, and all combinations were subjected to Swiss ADME (for Lipinski's rule, GIT Absorption P-GP substrate, similarity to lead), Pro-Tox II and Swiss doc. After combining all the data select some best peptides using AI techniques like Support Vector Machine, cosine similarity etc. and analyze the AI results (graphs, coding etc.) and filter out about 8 peptides and finally select one peptide among them.

Experimental Material and Methodology

Materials

The Configuration of the system in which the docking study was performed along with other online tools of drug design was running on 1.00GHz 1.19 GHz, system type 64-bit operating system, x64-based processor, Installed RAM: 8.00 GB (7.77 GB usable).

Methods

Forty-Five smaller chain dipeptides molecules Pro - Tyr, Pro - Trp, Pro – Phe, Pro – His, Pro– Lys, Pro – Leu, Pro – Arg, Pro – Asp, Pro – Cys, Pro – Met, Pro – Glu, Pro – Val, Pro – Ala, Phe – Ala, Phe – Trp, Phe – His, Phe – Cys, Phe – Val, Phe – Arg, Phe – Asp, Phe – Gly, Phe– Lys, Phe – Leu, Phe – Met, Phe – Val, Phe – Ser, Phe – Thr, Trp-Tyr, Trp – Thr, Trp – Gly, Trp – Asp, Trp – Ala, etc. were taken for the study and subjected to SWISS ADME online to check the drug likeliness property of Lipinski rule of five along with pharmacokinetics(What body does to the drug).

SWISS ADME - This is a free web-based tool that predicts pharmacokinetics, medicinal chemistry [8], lipophilicity, drug-likeness property, etc. Bioavailability of all compounds was tested against glycoprotein permeability (PGP) substrate along with drug metabolic profile against various cytochrome P450 inhibitions (CYP1 A2, CYP2C19, CYP2C9, CYP2D6 and CYP3A4). P-glycoproteins are efflux transporters in many organs that reduce drug absorption. The same set of compounds was subjected to online drug Way 2 [9], Way 2 drug PASS software to identify activity against a specific target with possible adverse and toxic effects and to check for toxicity using Pro-Tox II software. Pro-Tox II [10] Check toxicity according to 6 classes, free online tool Swiss Dock and CB Dock 2 [11] Used to dock selective dipeptides based on preliminary results obtained with the

target enzyme casein kinase 2, Toll-like receptor, which are sequentially obtained from zinc I. D database and PDB code (1JWH). The standard drug used for docking comparison with the tested dipeptides is Sorafenib and Lenvatinib. Visualization and detailed analysis of binding interactions in 3D and 2D was further performed using the UCSF Chimera tool and the Biovia Discovery studio visualizer. Support Vector Machine is a supervised machine learning type classifier that used labeled data to classify. In this study, a classification model has been developed using SVM. A total of 44 smaller chain dipeptides and 2 standard anti-cancer drugs were considered in this research work. The smaller chain dipeptides were converted into a format suitable for the SVM algorithm. To better classification we need to do preprocessing of the data set, so to preprocess the data set we need to make then a numerical value we have some value which is not numeric, so based on their nature we can assign them as a number value, we here just make a change 0 for low value and 1 for high value.

Gaussian Classifier is also a type of classifier like SVM for predictive modeling, however, SVM is often prone to unbiased classification datasets when you have more positive examples than negative ones, GPs don't usually suffer from this problem.

Cosine similarity is a measure of the cosine of angle between two vectors in a multidimensional space. so in our work what we did was firstly we converted all the Boolean features in o and 1 using label encoding and then we considered all the properties of a peptide sequence as multidimensional vector and then we did the same for standard drug and then calculated the cosine similarity between them and peptide with greater cosine similarity is more closer to the standard drug.

Results and Discussion

The Swiss ADME tool was used for all selected 45 smaller dipeptide chains with physicochemical properties (molecular weight, number of heavy atoms, number of aromatic heavy atoms, number of rotating bonds, number of hydrogen bond donors, number of hydrogen bond acceptors). , molar reactivity, total polar surface), lipophilicity (consensus log Po/w), water solubility (log S), all ADME properties (GI absorption, blood-brain barrier permeation, P-GP substrate, cytochrome P450 inhibitors CYP1A2, CYP2C19 , CYP2C9 , CYP2D6, CYP3A4, skin permeation (log KP), Lipinski's rule of drug similarity characteristics, bioavailability score, lead similarity, and synthetic availability as multiple parameters, and the results of all forty-five dipeptides are expressed in Table 1.

SWISS ADME results

Swiss ADME gives results about the molecular weight, Hydrogen Bond Donor and Acceptor, Molar refractivity, GIT Absorption, P-gp Substrate, and Lipinski, and on the behalf of important factors we eliminate the combinations that not show as a promising compound.

Table 1 represents the physicochemical properties of dipeptides.

Dipeptides	Molecular Weight	Hydrogen Bond Donor	Hydrogen Bond Acceptor	Molar Refractivity
Pro – Tyr	278	4	5	76.46
Pro - Trp	301	4	4	86.29
Pro – Phe	262	3	4	74.43
Pro – Ala	186	3	4	49.95
Pro – Gly	172	3	4	45.14
Pro – Val	214	3	4	59.56
Pro – Cys	218	3	4	57.88
Pro – Met	246	3	4	67.15
Pro – Leu	228	3	4	64.37
Pro – Thr	216	4	5	55.91
Pro – Ser	202	4	5	51.11
Pro – Lys	243	4	5	67.07
Pro – His	252	3	5	66.58
Pro – Asp	229	4	5	57.66
Pro – Arg	271	5	5	73.47
Pro - Gln	244	4	6	62.47

Table 2 represents Lipophilicity, medicinal chemistry, Pharmacokinetics & Drug likeness.

Dipeptides	Lipophilicity	GIT Absorption	P-GP Substrate	Lipinski	Lead Likeness
Pro – Tyr	0	High	No	Yes	Yes
Pro - Trp	0.54	High	No	Yes	Yes
Pro – Phe	0.33	High	No	Yes	Yes
Pro – Ala	-0.97	High	No	Yes	No
Pro – Gly	-1.17	High	No	Yes	No
Pro – Val	-0.29	High	No	Yes	No
Pro – Cys	-0.93	High	No	Yes	No

Pro – Met	-0.32	High	Yes	Yes	No
Pro – Leu	0.04	High	Yes	Yes	No
Pro – Thr	-1.35	High	Yes	Yes	No
Pro – Ser	-1.69	High	No	Yes	No
Pro – Lys	-0.88	High	Yes	Yes	No
Pro – His	-0.98	High	No	Yes	Yes
Pro – Asp	-2.03	Low	No	Yes	No
Pro – Arg	-1.73	Low	No	Yes	No
Pro - Gln	-1.2	High	No	Yes	No
Trp – Tyr	-0.49	Low	No	Yes	No
Trp – Gly	0.08	High	No	Yes	Yes
Trp – Val	0.85	High	No	Yes	Yes
Trp – Cys	0.15	Low	Yes	Yes	Yes
Trp – Ser	-0.47	High	No	Yes	Yes

After checking Lipinski, GIT absorption, P-gp substrate the elimination of combination take place on the basis of Low GIT absorption, Substrate for P-gp and not follow Lipinski rule like in Proline combination class Pro-Me, Pro-Leu, Pro-Thr, Pro-Lys are the substrate for P-gp so they are not considered as a promising combination. In the same ways the combination Pro-Asp, and Pro-Arg have Low GIT Absorption so not considered for good results. Also eliminate combinations in same manner from Tryptophan and Phenylalanine class. After doing SWISS ADME studies the next step check the toxicity class of our combination with the help of Pro-tox II software.

Table 3 represents toxicity class of dipeptides.

Dipeptides	**Class**
Pro – Tyr	5
Pro – Phe	5
Pro – Ala	5
Pro – Gly	5
Pro – Val	5
Pro – Cys	5
Pro – Ser	5
Pro – His	5
Pro – Gln	6
Trp – His	5
Phe – Ala	6
Phe –Cys	6

Phe – Gly	6
Phe – Glu	6
Phe – Lys	5
Phe –Leu	6
Phe – His	5
Phe – Tyr	6
Phe – Val	6
Phe – Ser	6
Phe – Thr	6
Phe –Met	6
Phe –Gln.	6

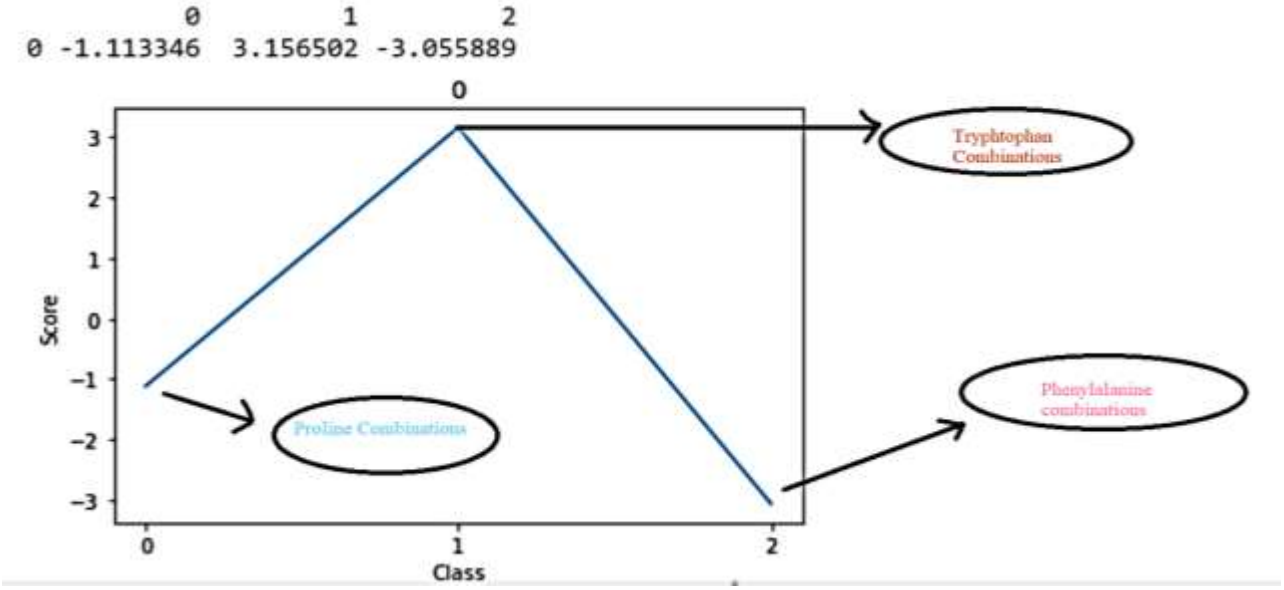
After checked the Toxicity class we consider those combinations as most prominent that lies in class 5 & 6 as they are more safer as in Tryptophan Class, the only combination that is considerable is Tryptophan-Histidine as they lie in class 5 and have higher GIT absorption as well as no substrate for P-gp, in Phenylalanine class almost all combinations are safe except few say one or two. Proline class has many safer combinations with high GIT absorption. All combinations were inactive against toxicity, which means none of the above combinations possess hepatotoxicity, carcinogenicity, or other toxicity. After checking toxicity, the receptor-ligand interaction is calculated by docking.

Table 4 represents the docking results of smaller chain dipeptides.

Dipeptides	**Receptors/Enzymes (Estimated ΔG (kcal/mol)) Standard Drug Value for CK2 is -9.37 &-8.85 and for TL-2 is -8.40 & -8.38**	
	Casein Kinase 2	**Toll-like Receptor 2**
Pro – Tyr	-7.70	-7.37
Pro – Phe	-8.10	-7.79
Pro – Ala	-6.74	-7.03
Pro – Gly	-6.77	-7.35
Pro – Val	-7.02	-7.95
Pro – Cys	-7.71	-7.25
Pro – Ser	-7.28	-7.06
Pro – His	-7.80	-7.54
Pro – Gln	-7.45	-7.34
Trp – His	-8.42	-7.59
Phe – Ala	-7.70	-7.37
Phe –Cys	-7.78	-8.49
Phe – Gly	-7.71	-7.17
Phe – Glu	-8.13	-9.98
Phe – Lys	-7.98	-7.72
Phe –Leu	-7.96	-7.67
Phe – His	-8.44	-7.95
Phe – Tyr	-9.06	-8.02
Phe – Val	-7.72	-7.39
Phe – Ser	-7.59	-7.36
Phe – Thr	-8.37	-7.98
Phe –Met	-8.12	-7.67
Phe-Gln	-8.30	-7.82

The docking result concluded that after eliminating some combinations almost 22 based on GIT absorption, P-gp substrate, Lipinski and Pro-Tox II result in the remaining dipeptides in which the docking result is considered In Proline Combinations for Casein Kinase 2:- Pro-Phe>Pro-His>Pro-Cys>Pro-Tyr and for Phenylalanine combination Phe-Tyr>Phe-His>Phe-Thr>Phe-Gln.

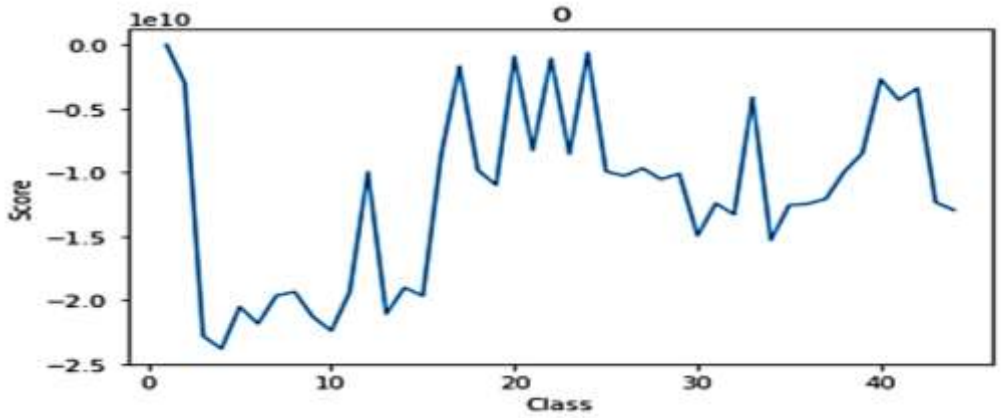
SVM result



Graph 1 represents the Support Vector Machine results

According to the SVM result, the most dipeptide combination that is similar to our standard drug lies in Tryptophan Class followed by Proline & Phenylalanine Class.

Gaussian Classifier result



Graph 2 represents the Gaussian classifier results.

Table 5 represents the score of all dipeptides in the Gaussian classifier.

Gaussian classifier score for individual peptides in the forms of e10									
0	-0.30	-2.28	-2. 38	-2.05	-2.18	-1.96	-1.93	-2.13	-2.24
-1.94	-0.99	-2.10	-1.90	-1.96	-0.86	-0.16	-0.98	-1.09	-0.09
-0.82	-0.10	-0.85	-0.06	-0.99	-1.02	-0.96	-1.05	-1.01	-1.49
-1.24	-1.33	-4.12	-1.53	-1.25	-1.24	-1.20	-0.99	-0.84	-0.27
-0.43	-0.34	-1.23	-1.29						

According to the Gaussian classifier, the top peptides that match with standard peptides are Pro-Trp>Trp-Thr>Trp-Ser>Trp-Ala>Trp-Gly>Phe-Val>Pro-Phe>Phe-Thr>Phe-Ser>Trp-Phe.

Cosine Similarity

Compounds like Sorafenib are Tryptophan-Aspartic Acid > Phenylalanine-Arginine Tryptophan-Phenylalanine > Tryptophan-Threonine

Compounds like Lenvatinib are Proline-Serine > Proline-Threonine > Proline-Leucine > Proline-Histidine.

Conclusion

After SWISS ADME on certain parameters, 23 dipeptides are filtered from 45 dipeptides, and we checked their toxicity and perform docking on them. After analyzing all result from docking [Man-made method], SVM, Gaussian classifier, and Cosine similarity.

Table 6 represents all the comparative results (Docking, SVM, Gaussian classifier, and Cosine similarity result) of smaller chain dipeptides.

Order	Docking Result	SVM Result	Gaussian Result	Cosine similarity result
1	Phenylalanine-Tyrosine [-9.06]	Tryptophan Class	Proline-Tryptophan	Proline-Serine
2	Phenylalanine-Histidine [-8.44]	Proline Class	Tryptophan-Threonine	Proline-Threonine
3	Phenylalanine-Threonine [-8.37]	Phenylalanine Class	Tryptophan-Serine	Proline-Leucine
4	Phenylalanine-Glutamine [-8.30]		Tryptophan-Alanine	Proline-Histidine
5	Proline-Phenylalanine [-8.10]		Tryptophan-Glycine	Tryptophan-Aspartic acid

As per Table 6, results concludes that on Docking Phe-Tyr was seen most promising lead dipeptide, on the other hand, AI approaches such as Gaussian show Pro-Trp is considered as the most promising & cosine similarity shows Pro-Ser.

Acknowledgement

The authors acknowledge and are grateful to Dr. (Col.) A. Garg, Director, and Dr. Manoj Goel who is the Joint Director of KIET Group of Institutions in overall support for carrying out of the work.

Conflict of Interest

All authors declare, that there is no competing interest.

References

[1]. Ozawa H, Hirayama A, Shoji F, Maruyama M, Suzuki K, Yamanaka-Okumura H, Tatano H, Morine Y, Soga T, Shimada M, Tomita M. Comprehensive Dipeptide Analysis Revealed Cancer-Specific Profile in the Liver of Patients with Hepatocellular Carcinoma and Hepatitis. Metabolites. 2020 Nov 1; 10(11):442. https://doi: 10.3390/metabo10110442.

[2]. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. Lancet. 2018 Mar 31; 391(10127):1301-1314. https://doi: 10.1016/S0140-6736(18)30010-2.

[3]. Ko CJ, Li CJ, Wu MY, Chu PY. Overexpression of LGR-5 as a Predictor of Poor Outcome in Patients with Hepatocellular Carcinoma. Int J Environ Res Public Health. 2019 May 23; 16(10):1836. https://doi: 10.3390/ijerph16101836.

[4]. Singh S, Kumar Sharma S, Mehrotra P, Bhatt P, Kaurav M. Blockchain technology for efficient data management in healthcare system: Opportunity, challenges and future perspectives. Mater Today [Internet]. 2022;62:5042–6. Available from:

http://dx.doi.org/10.1016/j.matpr.2022.04.998
[5]. Dodd-O J, Acevedo-Jake AM, Azizogli AR, Mulligan VK, Kumar VA. How to Design Peptides. Methods Mol Biol. 2023; 2597:187-216. https:// doi: 10.1007/978-1-0716-2835-5_15.
[6]. Ahamed S, Bhatt P, Sultanuddin SJ, Walia R, Haque MA, InayathAhamed SB. An Intelligent IoT enabled Health Care Surveillance using Machine Learning. In: 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI). IEEE; 2022.
[7]. Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Mol Divers. 2021 Aug; 25(3):1315-1360. https:// doi: 10.1007/s11030-021-10217-3.
[8]. Singh S, Pankaj B, Nagarajan K, P. Singh N, Bala V. Blockchain with cloud for handling healthcare data: A privacy-friendly platform. Mater Today [Internet]. 2022;62:5021–6. Available from: http://dx.doi.org/10.1016/j.matpr.2022.04.910
[9]. Zhang Y, Luo M, Wu P, Wu S, Lee TY, Bai C. Application of Computational Biology and Artificial Intelligence in Drug Design. Int J Mol Sci. 2022 Nov 5; 23 (21):13568. https://doi: 10.3390/ijms232113568.
[10]. Pittala S, Krelin Y, Shoshan-Barmatz V. Targeting Liver Cancer and Associated Pathologies in Mice with Mitochondrial VDAC1-Based Peptide. Neoplasia. 2018 Jun; 20(6):594-609. https://doi: 10.1016/j.neo.2018.02.012.