CLUSTERING-BASED APPROACHES FOR ELUCIDATING POSSIBLE THERAPEUTIC APPLICATIONS OF PHYTOCONSTITUENTS OF *Arnebia* AND *Dactylorhiza* FOUND IN THE HIMALAYAN REGION

Nitansh Sherpa[*], Khushi Kashyap[*], Manu Pant[**], Diksha Panwar[**], Bhasker Pant[***] , Kumud Pant[****1]
[*]Department of Biotechnology, Graphic Era (Deemed to be) University, Dehradun, Uttarakhand, India
[**]Department of Life Sciences, Graphic Era (Deemed to be) University, Dehradun, Uttarakhand, India.
[***]Department of Computer Science, Graphic Era (Deemed to be) University, Dehradun, Uttarakhand, India.
[****]Department of Biotechnology, Graphic Era (Deemed to be) University, Dehradun, Uttarakhand, India.

ABSTRACT
Health and quality of life can be improved by using medicinal plants. In today's scientific arena there is a huge need for integrating modern algorithms for proving worth of any traditional system of knowledge. Two such plants used traditionally to cure health related disorders are *Arnebia* and *Dactylorhiza. Arnebia* has medicinal properties for treating sore throats, fevers, and other ailments and *Dactylorhiza* is traditionally used for the treatment of dysentery, diarrhea, chronic fever, stomachaches, as well as wounds and burns. It is also used to increase regenerative fluid in debilitated women after childbirth and to treat general weakness. Very few experimental studies are done to prove the therapeutic importance of both these plants. In this research work, therefore Data Mining algorithm of Clustering implemented through WEKA has been used for proving the therapeutic value of both the plants.
The phytoconstituents of both the plants have been tested, using various attributes, to belong to a particular drug class and hence falling in the cluster to which these drugs belong. This kind of research work can open up multifarious ways of data analysis with merger of new techniques into traditional practices.

KEYWORDS: Medicinal plants, Medicinal properties, Clustering, Data Mining, Phytoconstituents, WEKA software

MSC: 62P10

RESUMEN
La salud y la calidad de vida se pueden mejorar mediante el uso de plantas medicinales. En el ámbito científico actual, existe una gran necesidad de integrar algoritmos modernos para probar el valor de cualquier sistema tradicional de conocimiento. Dos de estas plantas utilizadas tradicionalmente para curar trastornos relacionados con la salud son Arnebia y Dactylorhiza. Arnebia tiene propiedades medicinales para tratar dolores de garganta, fiebre y otras dolencias y Dactylorhiza se usa tradicionalmente para el tratamiento de disentería, diarrea, fiebre crónica, dolores de estómago, así como heridas y quemaduras. También se usa para aumentar el líquido regenerativo en mujeres debilitadas después del parto y para tratar la debilidad general. Se realizan muy pocos estudios experimentales para probar la importancia terapéutica de estas dos plantas. En este trabajo de investigación, por lo tanto, se ha utilizado el algoritmo de minería de datos de agrupamiento implementado a través de WEKA para probar el valor terapéutico de ambas plantas.
Los fito-constituyentes de ambas plantas han sido probados, usando varios atributos, para pertenecer a una clase de droga en particular y por lo tanto caer en el grupo al que pertenecen estas drogas. Este tipo de trabajo de investigación puede abrir múltiples formas de análisis de datos con la fusión de nuevas técnicas en las prácticas tradicionales.

PALABRAS CLAVE: plantas medicinales, propiedades medicinales, agrupamiento, minería de datos, fito-constituyentes, software WEKA

1. INTRODUCTION

There are many hidden jewels in the Himalayan region in the shape of lesser-known plants that have significant therapeutic value and can be extremely beneficial to humanity only if their potential is properly explored. These medicinal herbs have long played a significant role in indigenous people's lives. Two such lesser known ones

[1] pant.kumud@gmail.com (corresponding author)

are, *Arnebia* and its species and *Dactylorhiza* and its species traditionally used by the indigenous tribes of Himalayan region for healing, anti-pyretic and many more therapeutic effects as well as for general well being. Extracts from *Arnebia euchromatin*, in particular was also found to have anti-HIV activity by Kashiwada et al. [3]. Whereas the anti-inflammatory, antipyretic, anti-cancerous, neuropharmacological, and many more health effects of *Dactylorhiza* have also been reported by Wani et al. [8].

There is a need to understand the scientific basis of these traditional beliefs and practices. Although few laboratory experimentations for validating the therapeutic potential of these plants have been done, there is still a requirement for incorporation of modern day algorithms as a means of multifold analyses and hence reinforcement of beliefs with scientific values.

In this research work proof of medicinal potential of plants *Arnebia* and *Dactylorhiza* has been tried to be established via use of Data Mining algorithms, more specifically, k means clustering. This is a fresh endeavor to identify the therapeutic value of these plants, which could be extremely beneficial to humanity.

2. MATERIALS AND METHODS

The chemicals present in *Arnebia* and *Dactylorhiza* plants are compiled in table 1 by Shalu Devi Thankur [7], Anil Kumar et al. [4] and Dinesh Bhusal [2].Chemical descriptors for several phytochemicals present in these plants have been retrieved using Mordred software by Hirotomo Moriwaki et al. [5]. Graph of dissolution of various properties used for developing the classifier is shown in figure 1.

Table 1: Phytochemicals present in plants [4]

S. No.	Taxa	Family	Phytochemical constituents
1.	*Arnebia benthamii* (Wallich ex D. Don) I. M. Johnston.	Boraginaceae	shikonin, acetyl-shikonin, iso-butyryl-shikonin, β,β-di-methylacryl-shikonin, isovaleryl-shikonin, β-hydroxy-isovaleryl-shikonin, deoxy-shikonin, isobutyl-shikonin, arnebinone, arnebin-7, stigmasterol, etc. [4]
2.	*Dactylorhiza hatagirea*(D. Don).	Orchidaceae	ascorbic acid, butanedioic acid and dactylitis [2]

Clustering has been done using WEKA software by Witten IH and Frank E [9]. The clusters are depicted on x-axis and inertia on y-axis in figure 2.

In figure 3, the centroid is represented by a black dot. Each of the three sets of data is represented by a different color. The first set represents the number 0 in red, the second the number 1 in blue, and the third the number 2 in cyan. Cluster 3 seems to have widespread data.

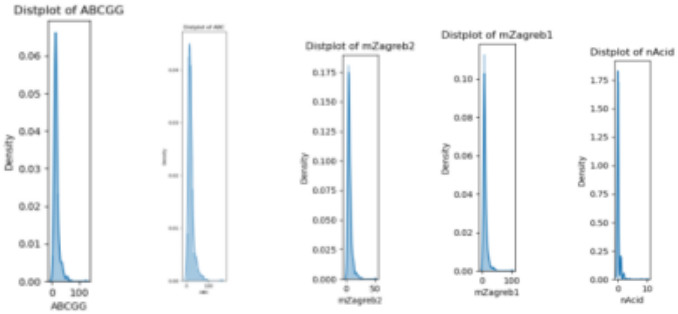
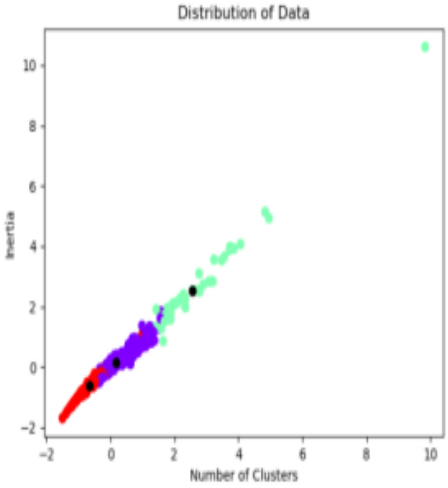


Figure 1: Graph of distribution of properties used for developing the clustering based classifier



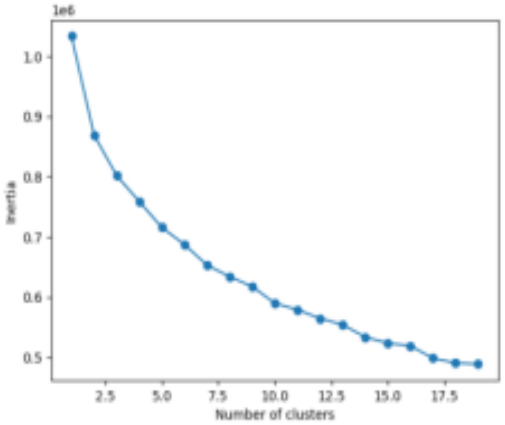


Figure 2: Representation of centroid of clusters. **Figure 3: Clusters depicted on the X-axis with inertia on the Y-axis**

The randomness of the cluster 3 can be seen in the expansion of the inertia values on Y-axis. For training the cluster generation algorithm, known drugs against various diseases like cancer, AIDS and malaria were used and phytoconstituents of the mentioned plants were tested on the train clusters.

The distance of the test sequences was compared to the centroid of the train sequences using Euclidean distances to determine the chance.

The distance of the phytochemicals in the collection based on few selected properties is shown in table 2. The phytochemicals discovered using the clustering method was found to have wide range of medicinal potential.

Table 2: Distance of the test sequence

ABC	**ABCGG**	**nAcid**	**mZagreb1 mZagreb2Class**	**Count**
count	675.000000	675.000000	675.000000 675.000000 675.000000	675.000000
mean	20.986379	16.940638	0.198519 10.0680566.006245	1.400000
std	14.046180	10.049067	0.6948297.204651 3.999833	1.587226
min 25% 50% 75%	0.000000 12.963281 17.860657 24.747527	0.000000 11.71428114.645916 19.002671	0.000000 0.0000000.000000 0.000000 6.3506943.8333330.0000008.055556 4.805556 0.00000011.559028 7076389	1.000000 1.000000 1.000000 1.000000
max	158.867557	123.444866	10.000000101.583333 49.722222	10.000000

Table 3: Phytoconstituents cluster and therapeutic property

Acetyl Shikonin cluster2 Anticancer molecules
Arabincluster2 Anticancer molecules
Deoxyshikonin cluster2Anticancer molecules
Dactylorhiza cluster1 Anti-HIV drug
Dactylose B cluster2 Anticancer molecule
Quercetin cluster2 Anticancer molecule
Dactylitis cluster 0 anti-malaria molecules

In table 3 the phytoconstituent falling in a particular cluster has been summarized. Similarly, in table 4 the drugs used for the study and cluster in which they fall are shown.

Table 4: Drugs used in the clustering method and their cluster

[1 1 1 1 0 1 1 0 1 0 0 1 0 0 0 1 2 0 1 0 1 1 0 0 0 1 0 1 1 2 1 0 1 0 1 0 0 0 1 0 1 1 1 0 2 1 1 2 0 0 0 0 0 0 1 0 1 0 0 0 0 1 1 1 1 1 0 0 1 2 0 0 1 0 1 1 0 0 1 1 0 1 0 0 0 0 1 2 1 1 1 1 0 2 1 2 1 0 2 0 1 1 0 1 0 1 2 2 2 1 1 1 1 1 1 1 1 1 1 0 0 0 1 0 0 0 1 1 1 0 1 1 1 0 1 0 2 1 1 1 1 0 0 0 1 2 1 1 2 0 1 0 0 1 2 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 0 0 0 0 1 1 1 1 2 1 0 1 1 0 0 0 0 1 1 1 0 2 1 0 1 0 2 0 0 1 1 0 0 1 2 0 1 0 2 0 0 0 1 0 0 0 2 1 1 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 2 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 2 2 0 0 1 1 0 1 1 1 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0 1 1 1 1 0 0 1 1 0 1 1 0 1 1 0 0 1 0 0 0 0 0 1 0 1 0 1 1 1 1 0 0 0 0 1 1 0 0 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 0 0 0 0 0 0 1 2 1 0 1 1 1 0 0 1 0 0 0 0 0 0 1 0 0 1 0 1 1 1 0 1 0 1 1 1 0 1 1 1 1 0 0 1 1 1 0 1 0 0 0 1 0 0 1 0 0 1 1 0 0 1 1 0 1 0 1 0 0 1 1 0 0 1 0 1 1 2 2 1 0 1 1 1 1 1 1 0 2 1 1 1 0 1 0 0 0 0 0 0 0 1 0 1 0 0 1 1 1 1 0 1 0 2 2 1 0 1 2 1 1 0 0 2 0 1 0 2 0 1 0 2 1 0 1 1 1 0 0 1 0 1 0 1 0 1 1 0 1 1 2 0 1 1 0 0 1 1 0 0 1 0 2 2 1 2 2 2 2 2 2 1 2 2 2 2 2 0 0 0 1 0 0 1 1 1 1 1 0 0 0 0 1 1 1 0 1 0 0 0 1 1 0 0 0 0 1 1 0 2 0 1 2 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
0 322
1 300
2 53

3. RESULTS AND DISCUSSION

During the study, it was found that Dactylorhin was previously indicated as an HIV inhibitor by Vijay Kumar Bhardwaj et al. [1]. In consistency with the previous results, Dactylose was found to belong to Cluster 1, which contains combination of anti-HIV chemicals. The matrix also displays how various substances are distributed in clusters. Compound 1 was discovered to be part of a cluster numbered 2 while compound Dactylitis was discovered to be part of a cluster numbered 0 referring to the first cluster and so on. A total of 322 train instances belonged to the first cluster (indicated by 0). The number of instances in the second cluster, represented by number 1, is 300, whereas the number of instances in the third cluster, represented by number 2, is 53.

We can therefore conclude that both plants have medicinal properties which are useful in curing various ailments as proved from clustering algorithms.

4. CONCLUSION

K-means clustering algorithm implemented through WEKA environment has successfully created clusters with various properties of known drugs against malaria, cancer and HIV. The phytoconstituents of the 2 plants under study on the basis of their molecular descriptors have been found to fall under various clusters. The clusters have been formed and trained with known medications against the three diseases viz. malaria, cancer and HIV and therefore the drug falling under a particular cluster can be said to possess the similar kind of properties. The current studies have shown some interesting results since phyto-constituent falling under a particular cluster are being tested in laboratories as deterrent to the same disease. The curative and therapeutic properties of plant components are being verified through this approach. Moreover, new therapeutic indications can also be proposed based on the similar concept. There are many diseases that can be used to make new drugs for epidemiological diseases if properly extracted. The antioxidant effects of plant roots also indicate that there is a promising future to the production of antioxidant compounds. It is the primary goal of this research to focus on the therapeutic possibilities for future research to be conducted in a way that the properties of the non-medicinal orchids can be taken into account. And this approach is considered for analysis/classification/application of the active ingredient(s).

RECEIVED: DECEMBER, 2022.

REVISED: MAY, 2023.

REFERENCES

[1]. BHARDWAJ, V. K., PUROHITA, R. and KUMAR, S. (2021): Himalayan bioactive molecules as potential entry inhibitors for the human immunodeficiency virus. **Food Chemistry,** 347, 128932**.**

[2]. BHUSAL, D. (2021): *Dactylorhiza hatagirea* (Panchaunle) and *Juglans regia* (Okhar): A review. **Asian Journal of Pharmacognosy,** 6, 27-33.

[3]. KASHIWADA, Y., NISHIZAWA, M., YAMAGISHI, T., TANAKA, T., NONAKA, G.-I. and COSENTINO L. M., et al. (1995): Anti-AIDS Agents, 18. sodium and potassium salts of caffeic acid tetramers from *Arnebia euchroma* as anti-HIV agents.**J. Nat. Prod.**, 58, 392–400.6.

[4]. KUMAR, A., SHASHNI, S., KUMAR, P., PANT, D., SINGH, A. and VERMA, R. K. (2021): Phytochemical constituents, distributions and traditional usages of *Arnebia euchroma*: A review. **Journal of Ethnopharmacology**, 271, 113896.

[5]. MORIWAKI, H., TIAN, Y-S., KAWASHITA, N. and TAKAGI, T. (2018): Mordred: a molecular descriptor calculator. **Journal of Cheminformatics,** 10:4. doi: 10.1186/s13321-018-0258-y.

[6]. PURI, B., and HALL, A. (1998): Phytochemical Dictionary: **A Handbook of Bioactive Compounds from Plants, Second** Edition (H. Baxter, J.B. Harborne, and G.P. Moss, Eds.) (2nd ed.). CRC Press, https://doi.org/10.4324/9780203483756.

[7]. THAKUR, S. D. (2019): Phytochemical constituents of some important medicinal plants from Dachigam National Park, Srinagar, Jammu and Kashmir. **The Pharma Innovation Journal**, 8, 68-71.

[8]. WANI, I. A., KUMAR, V., VERMA, S., TASLEEM JAN A. and RATHER, I. A. (2020): *Dactylorhiza hatagirea* (D. Don) Soo: A Critically Endangered Perennial Orchid from the North-West Himalayas. **Plants**, 9, 1644. https://doi.org/10.3390/plants9121644.

[9]. WITTEN, I. H. and FRANK, E. (2005): **Data Mining: Practical Machine Learning Tools and Techniques**, 2nd edition. Morgan Kaufmann, San Francisco.