Original Article

A predictive model for the early prognosis and characterization of asthma

***Pooja M R**
*Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru
Email: pooja.mr@vvce.ac.in

Abstract

An attempt to attain a good balance between optimal sensitivity and specificity of the predictive models in the case of sparse categorical data has been made in this paper by proposing a hybrid decision support system that integrates unsupervised and supervised learning methods at two different stages to explore the advantages of both. The system handles the categorical data without any encoding procedures involved. However, it requires one to use numerical categorical data in the place of labeled categorical data. The responses recorded in ACQ's are largely numerical data characterizing the presence or absence with 1 and 0 respectively. A single optional value indicative of a third response is often used representing an unknown response. The primary aim of the proposed work is to provide a platform for efficient outcome prediction that can assist in shared decision making. Shared decision making in the context of healthcare relates to a strategy where both the clinicians and patient themselves decide on par with the management of the disease.

Keywords: Sensitivity, Categorical, Encoding, Hybrid, Characterization

INTRODUCTION

Existing decision support systems and predictive models exhibit sub optimal performance with inadequate accuracy. Efforts to improve the same are ongoing and one such effort in this direction has been ours by adopting a solution that integrated both clustering and classification learning techniques in an effective way to obtain optimal prediction accuracies. The proposed solution is seen to outperform the traditional classifiers for prediction of disease outcome, as observed through an empirical analysis of the various standard learning classifiers.

The system thus developed offers a good sensitivity vs. specificity signifying that one can rely on the model developed to efficiently predict the disease. With this it can be integrated into any of the asthma tools that offer self-management questionnaires or any of the clinician offered tools that aim at performing preliminary screening for the disease. The evaluation of clinical assessment tools is basically done using two clusters, wherein there exists an almost equal number of subjects representing both categories, namely subjects with disease and without disease.

Access this article online	
Quick Response Code:	**Website:** www.pnrjournal.com **DOI:** 10.47750/pnr.2022.13.04.001

Address for correspondence: Pooja M R, Department of Computer Science and Engineering, Vidyavardhaka College of Engineering,
Mysuru,
Email: pooja.mr@vvce.ac.in

Received date: 15 August 2022 Accepted: 11 September, 2022
Published: 10 October, 2022

For reprints contact: pnrjournal@gmail.com

How to cite this article: Pooja M R, A predictive model for the early prognosis and characterization of asthma, J Pharm Negative Results 2022;13(4): 1-9.

Hence, we used the same class balanced data that was discussed earlier, as the input to the hybrid system for both of the datasets. This approach would validate the sensitivity and specificity calculated for the data under study, thereby assisting the clinicians to judge the ability of the system to distinguish those with the disease against those without the same. Sensitivity is the capacity of the model to recognize those with the disease as having the disease and specificity is its capacity to recognize correctly those who do not have disease [1,2,3]. However, the correctness of such an evaluation would always be done by referring to a gold standard[4,5,6]. The system developed, is however confined to categorical data as most of the asthma control questionnaires expect a response that is indicative either of a presence/absence of a disease symptom from the user and can be adopted when the decision making is to be done with respect to a set of questionnaires that follow the GOLD guidelines but, exclude numerical quantitative data that could possibly be gathered by clinical tests.

RELATED WORK

The traditional methods dealt with questionnaire response data from two different surveys covering information on distinct aspects of the disease characteristics that serve as the criteria for the assessment of the disease. The datasets were categorical in nature and classification task involved the prediction of a binary dependent variable using independent categorical responses. Some of the common challenges faced when dealing with categorical datasets obtained from various types of Asthma Control Questionnaires (ACQ) are presented below:

Few of the classes/levels of particular categorical variables rarely occur, and hardly impact the process of fitting a model. On the other hand, there are few classes among these variables which present themselves most frequently for most of the samples in the dataset, because of which they again fail to place a positive impact on the performance of the model, due to low variance. On the other hand if such variables are simply masked or skipped, we would be losing the information, following which we might not be able to interpret its contribution to the prediction process in any way[7,8,9]. The task is further impossible when a large number of variables represent the same characteristics. In the case of ISAAC cross sectional data recorded, most of the attribute values were recorded as ‘9” indicating a response that was neither positive nor negative for most of the questionnaires. However it is not feasible to simply skip such attributes which have an unknown response for most of the samples, as we might not be able to weigh the importance of such variables and their contributions in the prediction task. A good counter balance between sensitivity and specificity is difficult to achieve in situations such as the above involving sparse categorical data. The major reason being most of the traditional classifiers including the ensembles built using such classifiers relies on numerical data in order to perform effectively as they cannot operate on label data directly[10,11,12]. This poses a restriction on the efficient implementation of machine learning algorithms, thus necessitating the adoption of encoding procedures such as one hot encoding and integer encoding to do the required data transformation[7,8,9].

METHODOLOGY

Model Design

The primary purpose of the hybrid decision support system is to predict the disease outcome of asthma with a high degree of accuracy, by taking advantage of both supervised and unsupervised methods. The system is implemented in two successive stages, the first stage being the identification of target feature cluster that in turn identifies the relevant features using a modified version of Fuzzy C Means clustering, followed by a second stage which forms subject clusters such that the samples within each are characterized by a common phenotype. The prediction of disease outcome is done by fitting a binary classification tree on the input data which now includes categorical data indicative of identified subtypes of asthma generated from the previous stage involving subject clustering[13,14,15]. The identification of asthmatics and non-asthmatics is done with a high degree of specificity and sensitivity. Since the system was tailored for sparse categorical data we opted to test on ISAAC datasets for New Delhi and Neyveli which involved a considerable amount of underrepresentation for most of the attributes.

Each of the stages presented in the Figure 4.1 are explained below, the creation of the balanced data set is done and the same data obtained after class imbalance has been deployed for further validation of the hDSS. The system can be visualized as sequential as the functionality at every stage to a large extent depends on the results obtained by the previous stages[16,17,18].

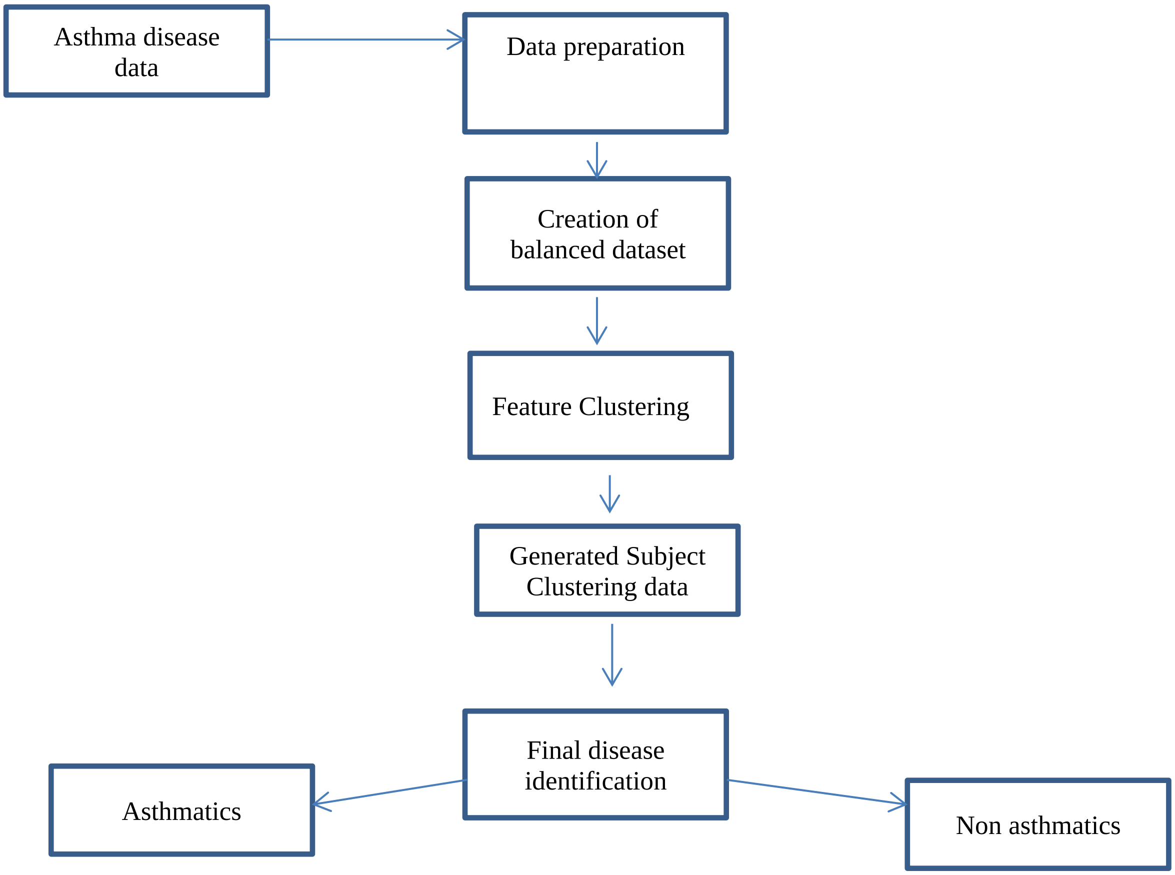

Figure 4.1. Predictive Model design

Data Acquisition and Preparation

The irrelevant attributes including form number and the version, country code and center id, school id, age group and date of recording information were eliminated manually. The "age in months" was discarded as redundant attribute as "age in years" is also present within the dataset. With manual elimination of these attributes we are left with 49 features out of a total of 58 features. It was observed from the data that some of the features have a "9" encoded for most of the samples in cases where a response was not recorded for such attributes. This necessities us to perform normalization. Further the samples indicating a value of "9" for asthma attribute were eliminated from the dataset. The features are initially normalized by applying z-score normalization, wherein all the attributes are scaled to accommodate a mean of zero and standard deviation of 1.The normalized dataset is now transposed to perform feature clustering[19,20].

Deducing optimal clusters

In the process of feature clustering, we applied subtractive clustering to infer the optimal number of clusters well before applying MFCM. As a one pass algorithm, subtractive clustering tries to estimate the number of clusters along with the cluster centers in the given data. This makes the computational complexity proportional to the size of the problem rather than its dimension. Though the cluster centers are necessarily not located at one of the actual data points, it gives a good approximation with reduction in the overall computational cost. As each of the point is a candidate for cluster center, a density measure for every point is defined using:

$$D_i = \sum_{j=1} \exp\left(\| X_i - X_j \|^2 / (r_\alpha/2)^2\right)$$

With this, a point having many neighboring instances will be regarded as highly dense. The point that has the highest density will be chosen as the first cluster center, following which the density measure of every point is revised using,

$$D_i = D_i - D_{c1} \exp\left(\| X_i - X_{c1} \|^2 / (r_b/2)^2\right)$$

After the revision, the next center will be the point having the greatest density value. This process repeats till an adequate and appropriate number of clusters are obtained.

Identifying Feature Clusters

FCM being an unsupervised learning technique basically necessitates a reasonable amount of supervision by the user to execute clustering and perform interpretation of its results. The choice of the number of clusters is ideally done by the user. We resorted to apply subtractive clustering with the range of cluster influence as 0.5 for all dimensions to the normalized features which suggested the feasibility of organizing the features into four clusters. Hence we choose to organize the features into 4 clusters by applying MFCM. With MFCM, we use a correlation based objective function. A different validity measure for proposed correlation based MFCM against the traditional Euclidean distance metric has been adopted[21,22,23].

The correlation-based FCM algorithm computes the fuzzy centroids cj and memberships uij at each iteration t as

$C(t)j=\sum Ni=1(u(t-1)ij)\ mxi\ \sum Ni = 1(u(t-1)\ ij)m\ ;j=1,2,\dots C$

And $u(t)ij=(\sum Ck=1[d(t)ijd(t)ik/]2(m-1))-1$

Where, C is the number of clusters. A fuzzy membership uij represents the membership of a feature i to a cluster cj, it satisfies a constraint $\sum Cj=1uij=1$.

A Pearson correlation coefficient is used in distance metrics dij which measure the difference between two features or between a feature and the cluster centroid. In most of the experimentations such as ours concerning FCM, the binary features may correlate positively or negatively. Correlated features should be clustered into one group regardless of their values being positive or negative. The distance metrics is defined as $dij,\ = 1-\ \rho2\ Xi,\ C\ j$, where $\rho Xi,C\ j$ is the Pearson correlation between a feature xi and a cluster cj. It would be 0 if they are highly correlated, either positively or negatively. Hence the steps to resolve the same would be as follows:

Initialize the cluster membership values, μij randomly

2. Compute the cluster centers: $C(t)j=\sum Ni=1(u(t-1)ij)mxi\sum Ni=1(u(t-1)ij)m;j=1,2,\dots C$

3. Update μij according to the following:

$u(t)\ ij =(\sum Ck=1[d(t)\ ij\ d(t)\ ij\ /]2(m-1))-1$

4. Calculate the objective function,

$Jm = \sum \mu ij\ d(xi,Cj) = \sum \mu ij[1-Corr(xi,Cj)2]$

5. Repeat steps 2–4 until *Jm* improves by less than a specified minimum threshold or until after a specified maximum number of iterations.

Table 4.1 depicts the feature clusters obtained for New Delhi data. It can be observed that the “asthma” attribute is located in the feature cluster 4 containing 9 features, hence we will be extracting the features "sex" , "whezev" , "exwhez12" , "cough12" , "pnoseev", "hfeverev" , "rashev ,"eczemaev" contributing to the reduced feature subset. These features will be used for the subsequent process of subject clustering followed by subject prediction. On similar lines, we extract 14 features that are identified with the asthma attribute within the cluster 3 in Table 4.2.

Table 4.1. Feature clustering results for New Delhi data

Cluster	Feature Clusters
Cluster 1	"whez12" "nwhez12" "awake12" "speech12" "pnose12" "ieyes12" "iactiv12" "rash12" "sitesev" "rclear12" "rawake12"
Cluster 2	"age" "rashage" "brthev" "brth12" "brth1m" "exbrthev" "exbrth12" "exbrth1m" "wwokenev" "wwoken12" "wwoken1m" "cwokenev" "cwoken12" "cwoken1m" "sabrthev" "sabrth12" "sabrth1m"
Cluster 3	"pnosejan" "pnosefeb" "pnosemar" "pnoseapr" "pnosemay" "pnosejun" "pnosejul" "pnoseaug" "pnosesep" "pnoseoct" "pnosenov" "pnosedec"
Cluster 4	"sex" "whezev" "exwhez12" "cough12" "pnoseev" "hfeverev" "rashev "eczemaev" "asthmaev"

Table 4.2 Feature clustering results for Neyveli data

Cluster	Feature Clusters
Cluster 1	"age" "rash12" "sitesev" "rashage" "rclear12" "rawake12"
Cluster 2	"whez12" "nwhez12" "awake12" "speech12" "brth12" "brth1m" "exbrth12" "exbrth1m" "wwoken12" "cwoken12" "cwoken1m" "sabrth12" "sabrth1m" "wwoken1m"
Cluster 3	"sex" "whezev" "exwhez12" "cough12" "hfeverev" "rashev" "eczemaev" "brthev" "exbrthev" "wwokenev" "cwokenev" "sabrthev" "pnoseev" "asthmaev
Cluster 4	"ieyes12" "iactiv12" "pnosenov" "pnoseoct" "pnosesep" "pnosejul" "pnosemar" "pnoseapr" "pnosemay" "pnosedec" "pnose12" "pnosejun" "pnosefeb" "pnosejan" "pnoseaug"

Identifying Subject Clustering and final disease identification

K-Means Clustering

K-Means clustering is an unsupervised learning algorithm that seeks to partition the input data into ‘K’ clusters by initially identifying ‘K’ random centroids. The data points are allocated to the nearest cluster, with the mean of the cluster being used to evaluate the distance between the data point and the cluster. The sum of squared distances is often used to calculate and estimate the nearness.

Step 1: Initialize the ‘K’ centroids to start with, and assign every data point to its nearest centroid. (Nearest is estimated using the data point that has the minimal sum of squared distance)

Step 2: Recompute the centroids as the mean of all the data assigned to the relevant clusters.

Step 3: Repeat step1 and step2 until there is no more significant change in the cluster centroids or a specified number of iterations are run on the data.

K-Means++ differs from the traditional K- Means clustering in that, the initial cluster centers rather than being chosen randomly are chosen such that they maximize the distance between the centroids using a procedure such as the below, in other words it attempts to choose the centroids such that they are far apart from each other.

Subject clustering was a preferred option as the asthma disease is known to exhibit several phenotypes and within-variation in the characteristics of individual phenotypes is always significant, with the disease being heterogonous as discussed in the section on introduction. The input data fed to subject clustering consists of 142*14 data matrix, representing 142 attributes represented with respect to 14 features. Similarly when the New Delhi data is fed as the input, a data matrix of size 220*9 will be utilized as the input representing 220 samples expressed with respect to 9 attributes identified in the feature cluster[24,25,26].

From the reduced feature set consisting of 142 samples expressed with respect to 14 variables, we eliminate thirty percent of the outliers by applying one class SVM fitted with a non-linear kernel function having a kernel co-efficient of 0.01. One class SVM is often deployed as an unsupervised learning technique to eliminate the errors within one class. It is quite possible that the procedure used to balance the class by adopting under sampling could carry with it some amount of outliers. Further, the asthmatic class category which is included as it is without any preliminary analysis on the amount of outliers it carries, also could be carrying a significant amount of noise within it. As a preventive step to recover from theses outliers we apply one class SVM on the balanced feature reduced data before performing subject clustering through K-means clustering.

The inliers retained are further used in the Subject Clustering by applying K-means clustering with random initialization for centroids. Rather than choosing a fixed number of clusters, we resorted to vary the number of clusters from 2 to 8 in order to determine an optimal choice. Silhouette scores were obtained by varying the number of clusters from 2 to 8 in order to decide the optimal number of clusters and are tabulated in the table below. Silhouette scores are strong indicators of quality of clusters as they try to contrast average distance to elements in the same cluster with the average distance to elements in other clusters. We opted to choose the number of subject clusters as two as it yielded higher Silhouette score compared to the others. The total number of iterations was 300 as the cluster centroids converged better only after this, with 100 re-runs carried out at every iteration [27,28].

The process to deduce the optimal number of clusters was carried out by repeating the process of K-Means clustering with K-Means++ as the initialization method. It was observed from this method too, that it is feasible to choose two clusters for subject clustering. On similar lines, we chose to have 8 clusters for the New Delhi dataset as it scored highest Silhouette value comparatively with both the initialization methods. The input data matrix of size 220*9 representing 220 samples expressed with respect to 9 attributes identified in the feature cluster was now used as the input to the system.

Prediction of Asthma outcome

The input to the prediction task contained the features constituting the reduced feature set along with the silhouette scores as one of the numerical input and the cluster labels as categorical inputs. An optimal binary tree obtained through experimental runs was chosen with the number of instances at leaves being restricted to 2. The minimum split size of the subsets was restricted to three and the maximal depth of the tree was confined to 10. Within the individual clusters, a binary classification tree is fitted using decision tree and the prediction results were recorded using the two testing schemes, 10-fold cross validation and Random sampling with a train/test split of 80 and 20 percent respectively. The accurate prediction of disease outcome is utmost important with a decent balance of sensitivity to specificity and the hybrid system proposed was able to do so, as verified by the comparative analysis of the results obtained by the same against the other classifiers discussed previously.

RESULTS AND DISCUSSION

Table 1: Silhouette scores for varying clusters

Initialization Method	Clusters	Silhouette score
Dataset	Neyveli data	
Random initialization	2	0.325
	3	0.236
	4	0.217
	5	0.220
	6	0.194
	7	0.204
	8	0.217
K-Means++	2	0.325
	3	0.236
	4	0.217
	5	0.216
	6	0.206
	7	0.219
	8	0.232
Dataset	New Delhi data	
Random initialization	2	0.384
	3	0.376
	4	0.409
	5	0.422
	6	0.442
	7	0.441
	8	0.464
K-Means++	2	0.384
	3	0.377
	4	0.405
	5	0.414
	6	0.446
	7	0.452
	8	0.461

Table 1 tabulates the Silhouette scores obtained using K-means clustering for clusters ranging from 2 to 8. While for the first dataset, a comparatively higher score of 0.325 is attained for two clusters, a higher score of 0.464 is obtained for eight clusters. As such, we opt to organize the subjects into two and eight clusters in the Neyveli and New Delhi data respectively.

The silhouette plots in figure 1 and 2 show the organization of the subjects into eight and two clusters for New Delhi and Neyveli data respectively.

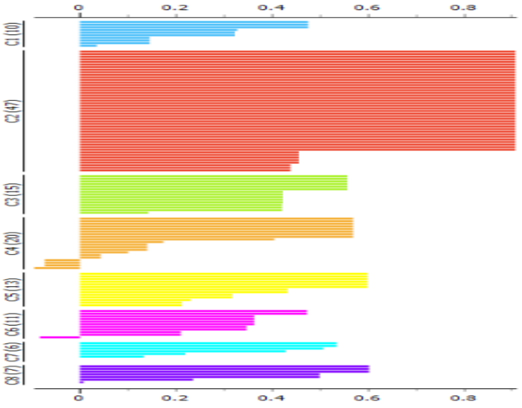

Figure 1: Silhouette plot for New Delhi data

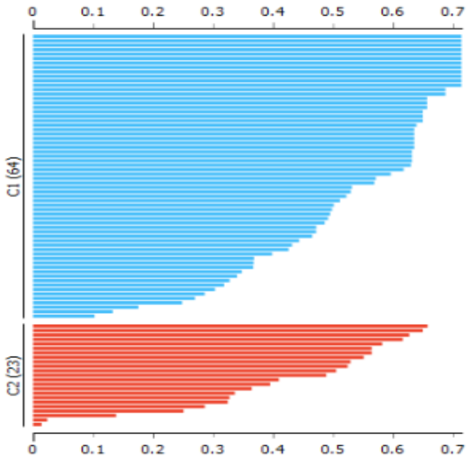

Figure 2: Silhouette plot for Neyveli

e distribution patterns of the two groups (asthmatics and non-asthmatics) within the individual clusters are well illustrated in figures 4.3 and 4.4.The data in blue represents asthmatics while 2 represent non-asthmatics. Table 4.4 shows the performance metrics evaluated for both the datasets using cross validation, stratified sampling and simple random sampling. The system yielded sensitivity and specificity of 95.3% vs. 93.6% as against 91.4% vs. 93.6% obtained using OGB technique in the case of New Delhi data. A sensitivity and specificity of 85.1% vs. 86.1% was obtained by hDSS with Neyveli data as compared to 84.5%vs. 82.8% achieved by OGB. Overall, it can be concluded that the hybrid system performs better than the ensembles for the reason that it tries to incorporate the efficacies of the clustering approach and its details into the learning model used for classification, resulting in a system with high predictive capabilities.

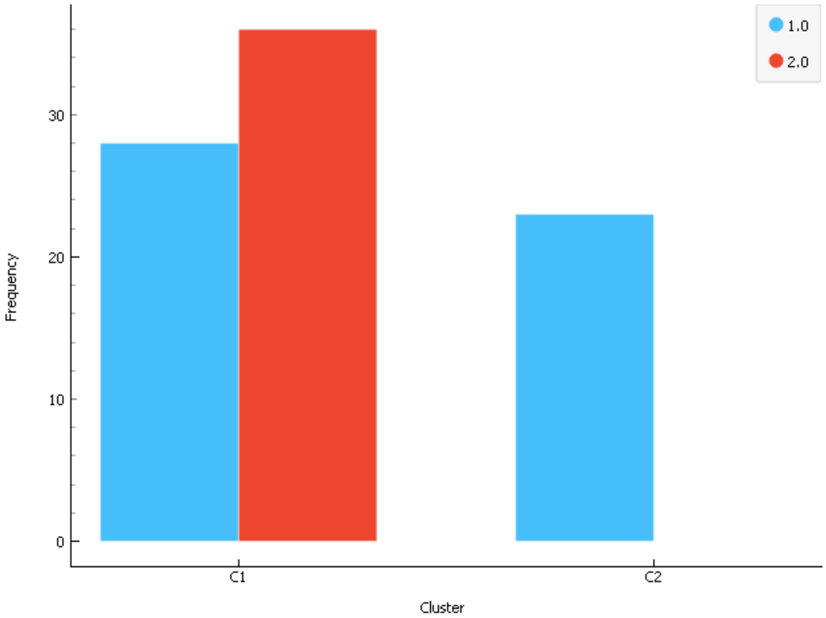

Figure 4.3: Distribution of the subjects within 2 clusters in Neyveli data

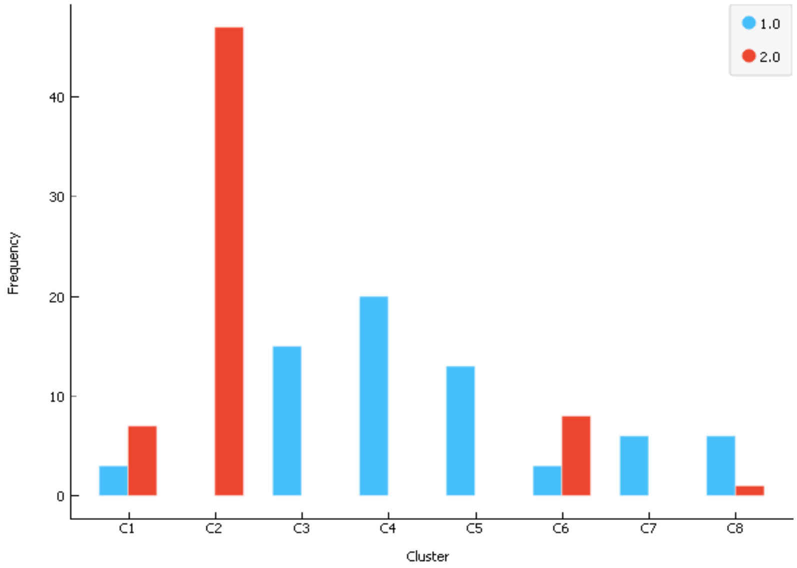

Figure 4.4: Distribution of the subjects within 2 clusters in Neyveli data

Table 4.4: Performance evaluation of the classifier using hDSS

Dataset	Testing Method	AUC	CA	F1	Precision	Recall
Neyveli	10-fold CV	0.855	0.851	0.851	0.854	0.851
	Stratified RS *	0.844	0.838	0.839	0.840	0.838
	RS*	0.845	0.834	0.835	0.835	0.834
New Delhi	10-fold CV	0.977	0.953	0.953	0.954	0.953
	Stratified RS*	0.972	0.945	0.945	0.945	0.945
	RS*	0.968	0.936	0.936	0.936	0.936

*RS- Random sampling

DISCUSSION

For the New Delhi data we had 8 individual clusters and the clusters were characterized as below:

Cluster 1 contained 10 subjects, out of which around three of them had asthma and were characterized by male children having persistent cough throughout the year. Cluster 2 contained non-asthmatics and as such we did not concentrate much on the inferences to be drawn. Cluster 3 completely contained asthmatics, which were characterized by cough along with wheeze and nose symptoms throughout the year and most of them were males. Cluster 4 also contained asthmatics and was dominated by male subjects with persistent wheezing. Cluster 5 contained asthmatics and had mostly males all of who had wheeze symptoms following exercise along with cough and wheezing in general. Cluster 6 contained a small percentage of asthmatics. Cluster 7 contained asthmatics, largely dominated by female subjects with nose and wheeze symptoms.

With the Neyveli Data, we had only two clusters, wherein one of the cluster completely contained asthmatics, largely dominated by males, had persistent wheeze symptoms along with the experience of wheeze symptom following exercise, breathing problems and problems in breathing following exercises and nose symptoms throughout the year. The other cluster was dominated by non-asthmatics and hence was not much targeted for our inferences.

CONCLUSION

The paper discusses the design and methodology involved in the development of hybrid decision support system to handle sparse categorical responses that is typical with most of the asthma response data. The work thrives at achieving objectives 1 and 3, by proposing a hybrid approach that integrates supervised and unsupervised learning techniques in an effective way to achieve optimal accuracy. The MFCM, a modified version of fuzzy c means clustering method adopted for the identification of feature clusters demonstrated good data dependency capabilities. Cluster characterization of asthma subject clusters is essential for identification of distinct asthma phenotypes to encourage personalized medical intervention, that is utmost important for the treatment of the disease. The system has effectively identified different asthma phenotypic groups with associated clinical risk factors, which is an important step taken towards advocating the right medications at early stages to overcome the negative impact of the disease. However, the clinical outcomes in this work were primarily done with respect to medical history alone. As such, in the subsequent works we would address the disease outcome prediction using lung assessment data obtained from spirometry in addition to the medical history data.

Authorship contribution

Pooja M R is the sole author who has contributed to the design and implementation of the research, leading to the analysis of the results, and writing of the manuscript.

Compliance with Ethical Standards

I Ethical approval – Not Applicable

II. Funding details (In case of Funding) – Not Applicable

III. Conflict of interest - Not Applicable

IV. Informed Consent – Not Applicable

REFERENCES

Chatzimichail, E., Paraskakis, E., & Rigas, A. (2013). Predicting Asthma Outcome Using Partial Least Square Regression and Artificial Neural Networks. Advances in Artificial Intelligence (16877470).

Manoharan, S. C., & Ramakrishnan, S. (2009). Prediction of forced expiratory volume in pulmonary function test using radial basis neural networks and k-means clustering. Journal of Medical Systems, 33(5), 347-351.

Wang, X., Wang, Z., Pengetnze, Y. M., Lachman, B. S., & Chowdhry, V. (2019). Deep Learning Models to Predict Pediatric Asthma Emergency Department Visits. arXiv preprint arXiv:1907.11195.

Tomita K, Nagao R, Touge H, Ikeuchi T, Sano H, Yamasaki A, Tohda Y. Deep learning facilitates the diagnosis of adult asthma. Allergology International. 2019;68(4):456-61.5.

Tobore, I., Li, J., Yuhang, L., Al-Handarish, Y., Kandwal, A., Nie, Z., & Wang, L. (2019). Deep learning intervention for health care challenges: some biomedical domain considerations. JMIR mHealth and uHealth, 7(8), e11966.

Loymans, R. J., Debray, T. P., Honkoop, P. J., Termeer, E. H., Snoeck-Stroband, J. B., Schermer, T. R., ... & Ter Riet, G. (2018). Exacerbations in adults with asthma: a systematic review and external validation of prediction models. The Journal of Allergy and Clinical Immunology: In Practice, 6(6), 1942-1952.

Deng, H., Urman, R., Gilliland, F. D., & Eckel, S. P. (2019). Understanding the importance of key risk factors in predicting chronic bronchitic symptoms using a machine learning approach. BMC medical research methodology, 19(1), 1-12.

Vial Dupuy, A., Amat, F., Pereira, B., Labbe, A., & Just, J. (2011). A simple tool to identify infants at high risk of mild to severe childhood asthma: the persistent asthma predictive score. Journal of Asthma, 48(10), 1015-1021.

Anastasiou, A., Kocsis, O., & Moustakas, K. (2018, July). Exploring machine learning for monitoring and predicting severe asthma exacerbations. In Proceedings of the 10th Hellenic Conference on Artificial Intelligence (pp. 1-6).

Messinger, A. I., Bui, N., Wagner, B. D., Szefler, S. J., Vu, T., & Deterding, R. R. (2019). Novel pediatric-automated respiratory score using physiologic data and machine learning in asthma. Pediatric pulmonology, 54(8), 1149-1155.

Spathis, D., & Vlamos, P. (2019). Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. Health informatics journal, 25(3), 811-827.

Sheshasaayee, A., & Prathiba, L. (2018). An Improvised Technique for the Diagnosis of Asthma Disease with the Categorization of Asthma Disease Level. In Information Systems Design and Intelligent Applications (pp. 985-994). Springer, Singapore.

Goto, T., Camargo Jr, C. A., Faridi, M. K., Yun, B. J., & Hasegawa, K. (2018). Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. The American journal of emergency medicine, 36(9), 1650-1654..

MR, P., & Pushpalatha, M. P. (2019). Clinical Respiratory Diseases and Care.

Pooja, M. R., & Pushpalatha, M. P. (2018). A neural network approach for risk assessment of asthma disease. Journal of Health Informatics & Management, 2(1), 1-6.

Pooja, M. R., & Pushpalatha, M. P. (2015, December). A hybrid decision support system for the identification of asthmatic subjects in a cross-sectional study. In 2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) (pp. 288-293). IEEE.

Pushpalatha, M. P., & Pooja, M. R. (2017, January). A predictive model for the effective prognosis of Asthma using Asthma severity indicators. In 2017 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). IEEE.

Pooja, M. R., & Pushpalatha, M. P. (2019). A predictive framework for the assessment of asthma control level. Int J Eng Adv Technol, 8, 239-45.

Badnjević, A., Gurbeta, L., Cifrek, M., & Marjanovic, D. (2016, May). Classification of asthma using artificial neural network. In 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 387-390). IEEE.

Salau, A.O., Pooja, M.R., Hasani, N.F., Braide, S.L. (2022). Model based risk assessment to evaluate lung functionality for early prognosis of asthma using neural network approach. Mathematical Modelling of Engineering Problems, Vol. 9, No. 4, pp. 1053-1060. https://doi.org/10.18280/mmep.090423

Xiang, Y., Ji, H., Zhou, Y., Li, F., Du, J., Rasmy, L., ... & Tao, C. (2020). Asthma exacerbation prediction and risk factor analysis based on a time-sensitive, attentive neural network: retrospective cohort study. Journal of medical Internet research, 22(7), e16981..

Haque, R., Ho, S. B., Chai, I., & Abdullah, A. (2021). Optimised deep neural network model to predict asthma exacerbation based on personalised weather triggers. F1000Research, 10.

Ma, X., Wu, Y., Zhang, L., Yuan, W., Yan, L., Fan, S., ... & Jia, W. (2020). Comparison and development of machine learning tools for the prediction of chronic obstructive pulmonary disease in the Chinese population. Journal of translational medicine, 18(1), 1-14.

Amaral, J. L., Sancho, A. G., Faria, A. C., Lopes, A. J., & Melo, P. L. (2020). Differential diagnosis of asthma and restrictive respiratory diseases by combining forced oscillation measurements, machine learning and neuro-fuzzy classifiers. Medical & Biological Engineering & Computing, 58(10), 2455-2473..

Amaral, J. L., Sancho, A. G., Faria, A. C., Lopes, A. J., & Melo, P. L. (2020). Differential diagnosis of asthma and restrictive respiratory diseases by combining forced oscillation measurements, machine learning and neuro-fuzzy classifiers. Medical & Biological Engineering & Computing, 58(10), 2455-2473.

Pooja, M. R. On Effective Use of Feature Engineering for Improving the Predictive Capability of Machine Learning Models. In Computational Intelligence and Data Sciences (pp. 53-62). CRC Press.

MR, P., & MP, P. (2020). Cluster Analysis to Characterize the Patterns of Complementary and Alternative Medicines Usage in Asthma Controls. The Open Public Health Journal, 13(1).