Using C4.5 Algorithm in Classification of Asthma in Children for Suggesting Best Possible Treatment

Ashish Jain and Swati Sharma
BSSS College, Bhopal, India b BSSS College, Bhopal, India
ashishkjain@bsssbhopal.edu.in, swati_98sharma@yahoo.co.in

Keywords: Data Mining, C4.5 algorithm, Pre-processing, Classification, Decision Tree.

Abstract: Millions of children worldwide suffer from asthma, and finding the best therapy is critical for treating the disease and improving the quality of life for those afflicted. Data mining is critical for detecting hidden patterns and trends in massive datasets, such as those used in healthcare. It has been used to identify and treat disorders including asthma. The C4.5 algorithm is a common decision tree technique that is employed in the proposed work to build a decision tree for selecting the optimal asthma medication in children. It employs three primary data mining steps: pre-processing, categorization, and decision tree. Finally, if the dependent variable matched the provided conditions, the results were gathered using a decision tree. Healthcare practitioners can make educated judgements by using data mining techniques.

1 INTRODUCTION

Data mining provides a mechanism for combining all methodologies naturally, allowing them to emphasise their strengths while concealing their limitations. As more data is generated in databases, classification analysis has emerged as a hot study area in data mining. Today, there are numerous classification algorithms accessible, including Decision Trees, Bayesian classification based on statistics, neural networks, and others.

However, it is important to apply such an algorithm which can deal with all types of symptoms, and thus helps in selection of the beat possible treatments for asthma based on symptoms entered into the database, for proper management of the asthma treatment very well in childhood by considering age groups such as 0-4, 4-8, and 8-12. According to the prior idea, the article provides a method based on a classification algorithm that makes use of a decision tree. Human processes are classified into two types in cognitive psychology: primary cognition and secondary cognition. Furthermore, the cognitive process utilises a variety of techniques. For intricate cases or objects, the most significant cognitive process of humans is to first classify the items and then further cognize each category in order to simplify the complicated things. Similarly, while building an application-specific algorithm for classification for asthma management based on the number of symptoms presented in the dataset, it is critical to be aware of the technique in order to simplify things. After categorizing asthma based on dependent and independent variables, there is other classification also which is intrinsic and extrinsic asthma, which is further subdivided into severe and general asthma. Input for classification is in the form of a .csv file, in which symptoms and their best possible Line of Treatment are saved and retrieved from the database. The section next describes a method of classification of data mining that can be used to find the most effective treatment among the many medications that are available. It also describes how to choose the selection variable. In particular, the classifier, testing the options and various attribute classes, and so on. In the final phase, the C4.5 algorithm is used to build a decision tree and choose the best asthma therapies for children.

2 CLASSIFICATION OF ASTHMA AS PER AGE GROUPS

Asthma is a worldwide disease, and its incidence is rising. It is predominantly a lung condition that manifests as the following symptoms:

- Reversible airway blockage, either naturally or by therapy
- Inflammation of the airways
- Hypersensitivity of the airways

Asthma is a disorder in which the airways in the lungs tighten and swell. It is frequent in children and teenagers. When a child has an asthma attack, his or her lungs do not receive enough oxygen to breathe, and they may cough or wheeze.
Based on age categories of 0-4, 4-8, and 8-12 years, the planned work will first identify asthma as intrinsic and extrinsic, then as severe and general. It employs a designed technique in which numerous symptoms like as coughing, wheezing, shortness of breath, chest tightness, and so on are recorded when a .csv file format is formed. The classification method is then applied to the entered data, and the classification is completed in conjunction with the decision tree. Having a diagnosed asthma, quantify the symptoms over a period of time as depicted in the Table 1.

Table 1: Asthma Symptoms.

Grades of Severity	Symptoms of Airflow	Night Time Obstruction Symptoms	Peak Expiratory Flow
Grade-4 Severe Persistent	Continuous with Limited Physical Activity	Frequent	>60% of Personal Best
Grade-3 Moderate Persistent	> Once a Day attack affects activity	> Once a Week	>60 % - <80%
Grade-2 Mild Persistent	> Once a Week but < Once a Day	> Twice a Month	> 80%
Grade-1 Mild Intermittent	< Once a Week Asymptomatic & Normal between attacks	< Twice a Month	> 80%

3 ATTRIBUTE SELECTION FOR ASTHMA DETECTION

The process of selecting attributes comprises searching through all potential attribute combinations in the data to discover which proportion of attributes works best for prediction. Variable selection is a challenging and critical topic in machine learning. In classification jobs, it can lead to greater accuracy or decreased computing costs.
This approach was used in this article to assign a value to each group of attributes. The C4.5 algorithm, which produces a decision tree, is used to determine the value of the dependent and independent variables. To build decision trees from a training data set, this approach employs the concept of information gain. It chooses symptoms as the data feature that best divides its sample set into subsets enriched in one of two classes.

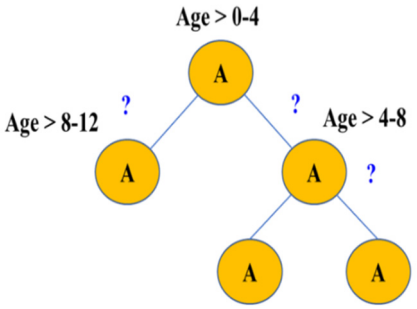

Figure 1: Classification of asthma as per Age group.

Its criteria is the normalized information gain (entropy difference) that occurs as a result of choosing an attribute for data splitting. The attribute with the largest normalized information gain is chosen to decide.

3.1 Classification Criteria Based Decision Tree

A decision tree is a predictive model that connects observations about an item to conclusions about its target value. It is also known as classification or regression trees.
The pseudocode algorithm based on the symptoms provided to the C4.5 algorithm for correct and accurate decision tree is as follows:

1. Check for basic symptoms for asthma
2. For each indefinite symptom of asthma as selection attribute **A**, find normalized information gain from splitting on **A**
3. Let **a_best** be the attribute with highest normalized information gain.
4. Create a decision node that splits on **a_best**
5. Recurse on the sub-lists obtained by splitting on **a_best** and add those nodes as $\mathbf{A_1}$, $\mathbf{A_2}$, $\mathbf{A_3}$, etc.
6. Those will be children of node **A** which has highest information gain.

For example:
Let the calculated information gain for 5 nodes will be, $A_1 = 0.98$, $A_2 = 0.86$, $A_3 = 0.5$, $A_4 = 0.74$, and $A_5 = 0.6$
Among above the node having highest information gain will be decision node like A_1. According to age group as a selection attribute, the proposed work got following results for the experiments done on symptoms table

Table 2: Classification Criteria.

Device	Age	For Preventer Regimes	For Acute Episodes (Home)	For Acute Episodes (Hospitals)
Metered Dose Inhaler (MDI)	Children > 10-12 yrs may learn appropriate usage. However, the Spacer is still recommended	For regimes incorporating *cromoglycate* or low dose inhaled steroid with or without long acting *beta2 agonist* (LA β_2 agonist)	May treat mild episodes	No role
MDI with Spacer	Suitable for all age groups. For smaller children (<3 yrs) attach a face mask	For all regimes, Recommended for Petients on medium to high dose inhaled steroid	Recommended for mild moderate episodes	Suitable for mild moderate episode
Dry Powder Inhaler (DPI)	May be used for children above 7-8 yrs of age	For regimes incorporating *cromoglycate* or low dose inhaled steroid with or without long acting LA β_2 agonist	May treat mild episodes	No role
Nebulizer	Suitable for all age groups	Do not recommend purchase	May be used	Recommended for Patients with severe episodes or ventilators

Table 3: Grading Table.

	First Choice	Other Options
Grade-4 Severe Persistent	Severe Persistent Medium to High Dose Inhaled Steroid + LABA (Formoterol Salmeterol) If needed add Oral Steroid	---
Grade-3 Moderate, Persistent	Low Dose Inhaled Steroid + LABA* or Medium dose Inhaled Steroid** If recurring Severe Exacerbation Medium dose in haled Steroid + LABA*	Low / Medium Dose Steroid + Leukotriene receptor antagonist / SR theophylline*
Grade-2 Mild Persistent	Low Dose Inhaled Steroid	Cromolyn LTRA, (Montelulcast) SR theophylline*
Grade-1 Mild Intermittent	No Daily Medication	---

4 CONCLUSION

In the twenty-first century, database and internet technology abilities have evolved fast. Meanwhile, a Management Information System and a Network Data Centre have been extensively used. Data access, data querying, and statistics all develop throughout time. However, the high layer's decision analysis and knowledge discovery are still immature, resulting in the phenomena of the "Information Explosion" and the "Knowledge Explosion". Data mining looks to be helping to solve these problems. Even though numerous sectors have investigated the classification algorithm problem, no algorithm can successfully handle a vast volume of data while also creating a decision tree.

Based on the findings of the studies, the suggested paper includes one datasheet that provides the best feasible treatment for better asthma management in children, which is the primary purpose of this study. (See Table III.) * For Children above 5 years only,

** For Children below 5 years

*** Evidence to date dose not supporting using a third long-term control medication added to inhaled corticosteroids and acting inhaled β agonists

in order to avoid using systemic corticosteroid therapy.
However, asthma can be very well managed from childhood as it is well known that asthma cannot be completely cured but can be controlled though such activities.

REFERENCES

Alexander I, Morton H. (2020), An Introduction to Neural Computing, 2nd edition Neural Networks at Pacific Northwest National Laboratory

Zhang Wx, Wu W Z (2018), Rough Set : Theory and Technique, Beijing Science Press

Paul E. Keller, Sherif Hashem (2022), A Novel Approach to Modelling and Diagnosing the Cardiovascular System, Proceedings of the workshop on environmental and energy applications of neural networks

Berikov, A. Litvinenko (2020), Methods for Statistical Data Analysis with Decision Trees, Novosibirsk Sobolev Institute of Mathematics, Methods for statistical data analysis with decision trees, 3(12-16)

Breiman, J. Friedman, R.A. Olshen, C.J. Stone (2020), Classification and Regression trees.

Yoos, H. L., & McMullen, A. (2022), Illness narratives of children with asthma. Pediatric Nursing, 22 (285-290)