



A clinical decision support system for diagnosis of Allergic Rhinitis based on intradermal skin tests



J. Jabez Christopher^a, H. Khanna Nehemiah^{a,*}, A. Kannan^b

^a Ramanujan Computing Centre, Anna University, Chennai 600025, Tamil Nadu, India

^b Department of Information Science and Technology, Anna University, Chennai 600025, Tamil Nadu, India

ARTICLE INFO

Article history:

Received 9 April 2015

Accepted 24 July 2015

Keywords:

Allergic Rhinitis

Allergens

Skin tests

Clinical decision support system

Rule base

ABSTRACT

Backgrounds and objectives: Allergic Rhinitis is a universal common disease, especially in populated cities and urban areas. Diagnosis and treatment of Allergic Rhinitis will improve the quality of life of allergic patients. Though skin tests remain the gold standard test for diagnosis of allergic disorders, clinical experts are required for accurate interpretation of test outcomes. This work presents a clinical decision support system (CDSS) to assist junior clinicians in the diagnosis of Allergic Rhinitis.

Methods: Intradermal Skin tests were performed on patients who had plausible allergic symptoms. Based on patient's history, 40 clinically relevant allergens were tested. 872 patients who had allergic symptoms were considered for this study. The rule based classification approach and the clinical test results were used to develop and validate the CDSS. Clinical relevance of the CDSS was compared with the Score for Allergic Rhinitis (SFAR). Tests were conducted for junior clinicians to assess their diagnostic capability in the absence of an expert.

Results: The class based Association rule generation approach provides a concise set of rules that is further validated by clinical experts. The interpretations of the experts are considered as the gold standard. The CDSS diagnoses the presence or absence of rhinitis with an accuracy of 88.31%. The allergy specialist and the junior clinicians prefer the rule based approach for its comprehensible knowledge model.

Conclusion: The Clinical Decision Support Systems with rule based classification approach assists junior doctors and clinicians in the diagnosis of Allergic Rhinitis to make reliable decisions based on the reports of intradermal skin tests.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Allergic diseases are the consequences of symptoms that occur due to an interaction between the environment and an individual's immune system, resulting in the discharge of histamine and other pro-inflammatory mediators. Allergic Rhinitis (AR) is a symptomatic disorder of the nose, induced after allergen exposure by an immunoglobulin E-mediated inflammation of the membranes lining the nose. It is characterised by nasal congestion, rhinorrhoea, sneezing, itching of nose and postnasal drainage [1]. Patients with Allergic Rhinitis may have nasal symptoms that affect their quality of life and their day-to-day activities [2]. A survey conducted in India by Kumar et al. reported that Allergic Rhinitis was prevalent in 26.1% of school children. When considered with the socio-economic status, 27.1% of the lower class people, 33.3% of the middle class people, and 28.6% of upper class

people living in urban areas were observed to have respiratory allergic diseases [3]. The characteristic symptoms of rhinitis were observed in 75% of children and 80% of asthmatic adults in India [4]. About 63.7% individuals affected with Allergic Rhinitis coexist with sinusitis. AR adds up to nearly 55% of all allergies present in India [5]. Furthermore there is very little research work conducted in the study of allergies and Allergic Rhinitis when compared to other areas in the medical domain. Literature and research, regarding the clinical features and symptoms of the patients with Allergic Rhinitis in India, is not exhaustive. Allergic diseases have not received the attention it deserves by both patients as well as clinicians [6].

Allergic diseases certainly require a unique gold standard test for detection and diagnosis. Works in literature use different tests for reporting an allergic disease. Gold standard tests that use only history findings of patients and physical examination are less sensitive and specific than those that incorporate specific immunoglobulin E (IgE) testing. The gold standard using a composite of the history, examination, food habits and specific IgE testing is the most commonly used method, but the methods of combining

* Corresponding author. Tel.: +91 044 22358013; mobile: +09381038339.

E-mail address: nehemiah@annauniv.edu (H. Khanna Nehemiah).

results has variations. Hence an appropriate method of assessment is necessary to assist clinicians in decision-making.

There are three gold standard tests which are used to diagnose Allergic Rhinitis, namely, Clinical criteria test, Composite test and Challenge test [7]. Clinical criteria gold standard test depends on the correlation between the patient's symptoms with clinical criteria set by expert opinion. Composite gold standard test use a combination of skin prick tests (SPT), in-vitro tests and nasal provocation tests. In challenge gold standard tests, the nasal passages are exposed directly to an allergen in increasing concentrations.

Diagnosis of Rhinitis requires a careful history analysis, well-targeted testing for allergic sensitization and physical examination. A prior knowledge about the aerobiology of the region in which the patient lives and patterns of allergen exposure is essential. Skin testing of a specific IgE is not easily accessible and cannot be interpreted precisely by primary care physicians in underserved areas [8]. Quality control measures and proper performance of skin testing are very important to produce correct results. Timely identification of allergens is vital as it may reduce the manifestation of symptoms. Allergens may vary depending on the age, potential allergen exposures, food habits and area of the country. Prick/puncture technique (percutaneous) or intradermal (intracutaneous) technique can be used for skin tests. Intradermal testing is more sensitive than prick/puncture testing [9]. To properly understand and interpret skin tests for specific IgE, it is essential to know which aeroallergens are present locally, are clinically important and have allergenic cross-reactivity with botanically related species [10]. Another limitation of SPT is that definitions of positive results of skin tests may vary from one physician to another. There are chances for inconsistent diagnosis. For example, many skin test studies compare the diameters of the wheals generated by allergens with the diameter of the wheal generated by the histamine control solution to classify test results as positive. Relatively unbiased test performance characteristics may enable junior clinicians to come out with confident decisions and suggestions. [7].

Though several methods exist for diagnosis and treatment of allergic diseases, a complete resolution of symptoms cannot be obtained. Computer-based predictive models are rarely used in diagnosis of allergic diseases [11]. Due to lack of independent unbiased allergy testing methods, this work proposes a computer-aided approach to assist clinicians in diagnosis and decision-making process. The development in computing and technology, has led to the use of Clinical Decision Support System (CDSS) in medical centres and hospitals for efficient and reliable diagnosis and prognosis. Furthermore CDSS are used for Dental education [12], Ventilator treatment [13], Cancer Diagnosis [14], Cancer Treatment [15], Diagnosis of heart diseases [16] and Heart Failure [17], Diagnosis of neuromuscular disorders [18] Respiratory pattern analysis [19] and many more. The CDSS in this work uses a rule based knowledge mining approach, for detection of Allergic Rhinitis from the results of the Intradermal Skin Tests.

The remainder of this paper is organised as follows: Section 2 presents the works in literature related to this work. Section 3 presents the framework of the Clinical Decision Support System along with the needed background definitions and notations. Section 4 provides information about the materials, dataset and the methods used in the experiment. Section 5 discusses the results followed by the conclusion and scope for future work.

2. Related work

Singh et al. have studied the effect of aeroallergens such as pollen, fungi, animal dander and dust in acting as triggers for

respiratory allergy. Specifically, they have surveyed the concentration of airborne pollen in different parts of India. Studies reveal that pollen types such as Casuarina, Parthenium, Spathodia, Eucalyptus, Peltophorum and Cyperaceae are dominant in southern parts of India [20]. Furthermore, Cassia, Ageratum, Salvadoria, Ricinus and Artemisia scoparia are considered to be important aeroallergens. Allergen-avoidance is an important measure in treating allergy. Their work does not provide any insight about diagnosis or treatment of aeroallergen-caused diseases.

Annesi et al. have developed a quantitative Score for Allergic Rhinitis (SFAR) ranging between 0 and 16 [21]. The score is based on eight criteria which are as follows: First, nasal symptoms such as running nose, blocked nose, second, accompaniment of symptoms with water eyes and itching, third, the season in which the symptoms occur, fourth, triggers of nasal symptoms such as inhalants, fifth, perceived allergic status, sixth, previous medical diagnosis of allergy, seventh, allergy tests reported to be positive and finally, family history. Data were obtained from 269 out-patients during a period of 30 days in 1997 at ENT clinics of four hospitals in France. The score was also validated with doctor's diagnosis and psychometric tests. SFAR achieved 74% sensitivity and 83% specificity. The score is based on Skin Prick Tests and ImmunoglobulinE levels and hence it cannot be used independently for diagnosis.

Chae et al. have developed an asthma diagnosis system [22]. The authors have compared the diagnostic efficiency of three supervised machine learning techniques namely neural network (NN), case-based reasoning (CBR) and discriminant analysis. Data were collected from 294 patients at Bronchial Asthma clinics at Yonsei University Severance hospital. The neural network is trained using a gradient descent based back-propagation algorithm. Twenty one characteristics of the patient and the results of the SPT constituted the input to the NN. Five hidden layer nodes with 20,000 training cycles produced optimal results. Compared to the diagnostic results of a doctor, the neural network achieves 92% accuracy. The CBR model was implemented using an expert system tool, ART-IM (Automated Reasoning Tool for Information Management). It achieved an overall accuracy of 84%. In discriminant analysis knowledge model, the authors have analysed the patient characteristics based on 14 questionnaire items, lab tests and SPT. The characteristics were grouped into three homogeneous factors. The discriminant function predicted the asthma status. The system achieved an overall accuracy of 90%. The authors have worked on a similar system for Allergic Rhinitis over 444 patients at the Otolaryngology Department of Inje University Paik Hospital. The system achieved 76%, 62% and 78% for NN, CBR and discriminant analysis respectively.

3. System framework

The Clinical Decision Support System uses a typical rule based knowledge mining approach to analyse the results (data) from the SPT and diagnose the presence/absence of AR. This section presents the working of the system with a detailed description of the framework and the mathematical models.

3.1. Definitions and representations

Let D represent a dataset consisting of rows and columns. D consists of k distinct attributes $A_1, A_2, \dots, A_k$ and a class attribute C . Each attribute has a definite domain $A_1 = \{a_{11}, a_{12}, \dots, a_{1m}\}$, $A_2 = \{a_{21}, a_{22}, \dots, a_{2m}\}, \dots, A_k = \{a_{k1}, a_{k2}, \dots, a_{km}\}$ and $C = \{c_1, c_2, \dots, c_{mc}\}$ where $c_1, c_2, \dots, c_{mc}$ are class labels. Each row (sample/instance), r , in D corresponds to a record and can be described as $\langle A_1 = a_{1,1,m} \rangle, \langle A_2 = a_{2,1,m} \rangle, \dots, \langle A_k = a_{k,1,m} \rangle, \langle C = c_{1,m} \rangle$.

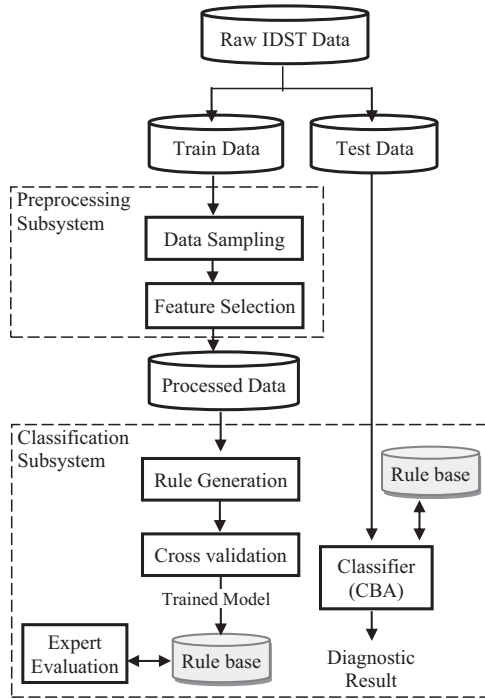


Fig. 1. System framework.

Number of rows N_r in D represent the number of instances in D , which may be also represented as $|D|$.

A classification rule, r , is of the form Antecedent $\rightarrow$ Consequent ($X \rightarrow Y$)

$$X \ni x_1, x_2, x_3, \dots, x_k$$

where $x_1 = \langle A_1 = a_{1,1..m} \rangle$, $x_2 = \langle A_2 = a_{2,1..m} \rangle$, ..., $x_k = \langle A_k = a_{k,1..m} \rangle$ and $Y \ni y_c$ where $y_c = \langle C = c_{1..m} \rangle$.

A ruleset R is a collection of rules, $R \ni \{r_1, r_2, \dots, r_n\}$.

Support of a classification rule r , $r: X \rightarrow Y$, refers to the per cent of cases in D that contain r [30].

$$\text{Support}(r) = \frac{\text{Count}(r)}{|D|} = \text{Probability}(X \rightarrow Y)$$

Confidence of the rule r refers to the ratio between the number of cases in $|D|$ that belong to r to the number of cases whose Antecedent X is the same as that of r

$$\text{For } r: X \rightarrow Y \text{ Confidence}(r) = \frac{\text{Count}(r)}{\text{Count}(X)} = \text{Probability}(X \rightarrow Y/X)$$

Rules below the minimum support (*minsup*) and minimum confidence (*minconf*) threshold are pruned.

3.2. System framework

The framework of the clinical decision support system is shown in Fig. 1. The major components of the CDSS system are the preprocessing subsystem and classification subsystem. The results of the Intra dermal skin test constitute the raw data. The data is partitioned into two mutually exclusive partitions called the train data and test data, using a holdout approach [23]. Two-third of the data is designated as train data (training set) and the remaining is the test data (testing set).

3.2.1. Preprocessing subsystem

The data is preprocessed in order to improve the quality of the data. The preprocessing subsystem performs data sampling and

feature (attribute) selection. A dataset is imbalanced if the classification categories are not approximately equally represented [24]. In general, a natural distribution is often not the best distribution for a learning algorithm (inducer) that is used to build a classifier. In this work, the train data has 581 samples out of which only 210 belong to the positive class. Since the class distribution in the raw IDST dataset is imbalanced, an oversampling technique is used to produce a balanced class distribution.

Input: Raw IDST train data

$B_{\text{SAMPLES}} = \{\Phi\}$ //set containing borderline samples of minority class

Process: Borderline Synthetic Minority Oversampling Technique [25]

Step 1: **For** each sample ' r ' in minority class, calculate m nearest neighbours ($m=5$).

begin

Step 2: **If** all m neighbours belong to majority class **then** r is considered as noise

Step 3: **If** r 's majority nearest neighbours is larger than the minority ones **then** $B_{\text{SAMPLES}} \leftarrow r$

end

Step 4: **For** each r in set B_{SAMPLES} **begin**

Step 5: Compute the m nearest neighbours

Step 6: Choose a neighbour at random, and compute the difference (*diff*)

Step 7: Generate a random value (*gap*) between 0 and 1

Step 8: Synthetic sample = Actual sample + $\text{gap} \times \text{diff}$
REPEAT Steps 5 to Step 8 for each attribute A in a sample r

end

Output: Dataset with a balanced class distribution

The balanced data set, is subjected to feature selection (attribute reduction). Table 1 presents the complete feature set (attributes). It consists of allergens (inhalants, contactants, ingestants), allergic symptoms and other information regarding a patient. Attribute indices 1–80 are allergens, 81–88 are the allergic symptoms and 89–91 are information regarding the patient. The values for the allergens are the results of the skin test. The allergic symptoms are binary valued (yes/no) attributes. The attribute 'Family History' is also a binary valued attribute which states the presence/absence of allergic symptoms among the members of the patient's family.

Three approaches were used for attribute evaluation and selection. First, the Information Gain ($I_A(D)$) [26] of all attributes was computed. Second, all the attributes were evaluated and ranked based on the worth of an attribute by measuring the correlation (Pearson's) with the class attribute. Third, the attributes were validated by the allergy specialist and suggestions were incorporated. Expert's opinion on the choice of attribute was more valid than gain-based and correlation selection procedure. Hence the combination of these three approaches gives an optimal subset of attributes. Out of 91 attributes (features), 40 were selected. The remaining allergens, generally, are not significant triggers to allergic symptoms among most patients who participated in the study.

Input: Balanced Dataset

Process:

A dataset D has $|D|$ samples belonging to ' m ' distinct classes. The expected reduction in the information requirement caused by knowing the value of an attribute A is given by Eq. (1)

$$\text{Gain}(A) = I(D) - I_A(D) \quad (1)$$

Table 1

List of allergens, allergic symptoms and patient details.

Inhalants, Contactants and Ingestants (ALLERGENS)							
1	Housedust	21	Fish1 ^a	41	Avaraikai (Broad beans)	61	Gram ^a
2	Cotondust	22	Fish2 ^a	42	Kovaikai (Coccinia grandis)	62	Channa
3	Aspergillus	23	Crab	43	Kothavarai (Cluster beans)	63	Dhal
4	Pollen	24	Prawns	44	Lady's Finger	64	Maida
5	Parthenium	25	Shark	45	Malli (Coriander)	65	Oats
6	Cockroach	26	Gourds ^a	46	Mango	66	Ragi
7	Cat Dander	27	Banana	47	Mushroom	67	Rice
8	Dog Fur	28	Beans	48	Nuckol (Brassica oleracea)	68	Wheat
9	Road Dust	29	Beet Root	49	Onion	69	Coconut
10	Old Paper Dust	30	Brinjal	50	Peas	70	Oil ^a
11	P.S Dust	31	Cabbage	51	Potroot	71	Garlic
12	Milk(P)	32	Capsicum	52	Paneer (Farmer cheese,)	72	Ginger
13	Milk(C)	33	Chillie	53	Potato	73	Pepper
14	Curd	34	Cauliflower	54	Pumpkin	74	Tamarind
15	Coffee	35	Carrot	55	Pudina (Mentha spicata)	75	Aginomoto
16	Tea	36	Radish	56	Chow chow (<i>Chayota edulis</i>)	76	Spices ^a
17	Beef	37	Corn	57	Tomato	77	Coco
18	Chicken	38	Cucumber	58	Tondaikai (Trichosanthes dioica)	78	Horlicks
19	Mutton	39	Drumstick	59	Plantain Stem	79	Boost
20	Egg	40	Greens ^a	60	Yams	80	Nuts ^a
ALLERGIC SYMPTOMS and PHYSICAL ATTRIBUTES							
81	Running Nose	85	Nasal Blocks	89	Age		
82	Sneeze	86	Headache	90	Gender		
83	Cough	87	Itching	91	Family History		
84	Wheezing	88	Rashes	92	Class		

^a Customised based on patient's history.

$$I_A(D) = \sum_{i=1}^{am} \frac{|D_i|}{|D|} \times I(D_i) \quad (2)$$

$$I(D) = - \sum_{i=1}^{mc} p_i \log_2(p_i) \quad (3)$$

where p_i is the probability that an instance belongs to class C_i is estimated by

$$p_i = \frac{|C_i, D|}{|D|}$$

Attribute 'A' has distinct a values, $1 \leq a \leq am$, and the Class Attribute 'C' has mc classes.

The Pearson's correlation coefficient (ρ) for an attribute A (ρ_A) is computed to validate the results of the Information Gain based Attribute evaluation.

$$\rho_A = \frac{1}{|D|-1} \left(\frac{a - \bar{A}}{SD_A} \times \frac{c - \bar{C}}{SD_C} \right)$$

where $\bar{A}$ and $\bar{C}$ is the mean of Attribute A and C respectively SD is the standard deviation.

However, Pearson correlation coefficient [27] estimated for two categorical variables will return the phi coefficient, ϕ which is computed as follows:

$$\rho_A = \phi^2 = \frac{\chi^2}{|D|}$$

where χ^2 is the Pearson's Chi-Squared statistic [28].

Based on the suggestions of the allergy specialist and the inferences from the above tests an optimal subset of features is selected.

Output: Processed dataset with optimal set of attributes.

The information gain and the correlation coefficient values for all attributes are presented in Appendix I.

3.2.2. Classification subsystem

The classification subsystem consists of rule generation, cross validation, expert evaluation, a rule base and a classifier. The CBA-

RG (Class-Based Association Rule Generation) algorithm is used for rule generation [29].

3.2.3. Rule generation

Input: Training dataset

Process:

Step 1: Generate candidate rules using Apriori approach [30]

FOR EACH Rule r

BEGIN

Step 2: Evaluate $Confidence(r)$

Step 3: If $Confidence(r) > minconf$

Then $R \leftarrow r$

END

Step 2: Order the rules based on Support and confidence

Interestingness measures.

Rule r_i precedes r_j if $Confidence(r_i) > Confidence(r_j)$

else if $Confidence(r_i) = Confidence(r_j)$ then r_i precedes

r_j if $Support(r_i) > Support(r_j)$

else if $Confidence(r_i) = Confidence(r_j)$ AND $Support$

$(r_i) = Support(r_j)$ then the rules are placed in the generated order.

Output: Ruleset R

Fig. 2 presents the contents of the rule base before expert validation

Low threshold values resulted in generation of more than forty thousand rules; hence the $minsup$ and $minconf$ values were set to 0.5 and 0.9 respectively. However, some potential rules that were filtered were incorporated on expert validation. Finally the rules are stored in the rule base for classifying instances from the test dataset.

3.2.4. Cross-validation

Cross-validation (CV) with ' k ' folds is a technique whereby the dataset D , is randomly split into k folds of approximately equal size. The classifier (model) is trained and tested k times. Each time ($k-1$) folds are used for training and the remaining one fold is used for testing. In classification, k -fold cross-validation is the best method to

```

1. running nose=no wheezing=no ==> Rhinitis=negative
2. running nose=no cough=no ==> Rhinitis=negative
3. cottondust=R sneeze=no ==> Rhinitis=negative
4. shark=0 running nose=no cough=no ==> Rhinitis=negative
5. pumkin=0 running nose=no sneeze=no ==> Rhinitis=negative
6. sneeze=no headache=no ==> Rhinitis=negative
7. kovaikai=0 running nose=no sneeze=no ==> Rhinitis=negative
8. running nose=no sneeze=no headache=no ==> Rhinitis=negative
9. shark=0 pumkin=0 running nose=no sneeze=no ==> Rhinitis=negative
10. running nose=no sneeze=no ==> Rhinitis=negative
11. shark=0 sneeze=no headache=no ==> Rhinitis=negative
12. shark=0 kovaikai=0 running nose=no sneeze=no ==> Rhinitis=negative
13. shark=0 running nose=no sneeze=no ==> Rhinitis=negative
14. pumkin=0 sneeze=no ==> Rhinitis=negative
15. kovaikai=0 sneeze=no ==> Rhinitis=negative
16. shark=0 pumkin=0 sneeze=no ==> Rhinitis=negative
17. sneeze=no ==> Rhinitis=negative
18. shark=0 kovaikai=0 sneeze=no ==> Rhinitis=negative
19. shark=0 sneeze=no ==> Rhinitis=negative
20. kovaikai=0 running nose=no headache=no ==> Rhinitis=negative
21. cottondust=R running nose=no headache=no ==> Rhinitis=negative
22. shark=0 cough=no ==> Rhinitis=negative
23. cough=no ==> Rhinitis=negative
24. running nose=no headache=no ==> Rhinitis=negative
25. shark=0 running nose=no headache=no ==> Rhinitis=negative

Fallout Rule: Rhinitis=positive

```

Fig. 2. Contents of the rule base after rule generation.

use for validating and selecting a classifier [31]. Associative classifier (CBA), Decision tree classifier (C4.5), Support Vector Machine (SVM), Multi-layer perceptron classifier (MLP), Naïve Bayes classifier (NB) and k-nearest neighbour classifier (kNN) are validated. In order to make the test unbiased, cross validation is applied over the same features and same partitions. The samples in each partition remain the same when each fold is iteratively tested. However, different runs had different samples in the folds in order to avoid the variations and perturbations that may exist due to cross validation. 10 independent runs of 10-fold cross-validation were performed. The complete results of cross validations are presented in [Appendix II](#).

3.2.5. Expert validation

The best rules extracted by the system (based on support and confidence measures) are validated. A panel of experts were involved in the validation process. The rules and results after processing the data were presented to them. Their combined suggestions were then incorporated. Clinical judgement is far more comprehensive than pure mathematics and machine learning. There may be additional subconscious factors which are overlooked by the model. Though the development of the decision support system is automated using data mining techniques, it is not possible to represent the complete knowledge and experience of an expert [32]. Hence the expert makes minor changes based on his experience. Furthermore, the CDSS is intended to be used by clinicians in southern parts of India. Hence these adjustments will not lead to over-fitting but makes the CDSS more apt for the prevailing situation.

3.2.6. Classifier

The classifier interacts with the rule base, and fetches the best matching rule for a query instance.

Input: Test data

Process:

Step 1: Match the attribute values of the test data (query

Instance) with the rules in the rule base.

Step 2: The class label of the most fitting rule that is of higher precedence is class for the test sample.

Step 3: If there is no rule to match the sample (test query), the fallout rule (Default rule) [23] is used for classification.

Output: Diagnostic result

4. Materials and methods

The clinical history of the patient is obtained which include name, age, gender (physical attributes) and allergic symptoms. A customised list of allergens was tested based on the lifestyle and food habits of the patients. The complete list of allergens based on the food habits and environmental conditions of the people of Chennai, South India, is presented in [Table 1](#).

Intradermal skin tests were performed following established methods using the allergen extracts, negative controls (saline) and positive controls (histamine). The upper half of the volar surface of the forearm was selected for the test. About 0.01 ml of the allergen is injected into the epidermis using sterile, disposable, plastic 1 ml tuberculin syringe. Patients were asked to stop taking antihistamines and anti-allergic drugs and medications. [Table 2](#) lists the medication to be avoided before allergy testing. Consecutive readings, on an hourly basis, were taken. A positive reaction is depicted by a wheal and a flare reaction [33]. A negative response (NR) to a skin test usually indicates that the patient is not sensitive to that allergen.

Intradermal Skin Test (IDST) data were collected from 872 patients who visited the Good Samaritan Kilpauk Lab and Allergy Testing Centre, Chennai, Tamilnadu, India during October and December 2012, March and April 2013. The patients were referred by ENT Surgeons and general physicians because of nasal problems or other plausible allergic symptoms. 413 males and 459 females, of all age groups, were included in the study.

4.1. Experimental settings

The Association Rule based classification approach used in this framework is compared with other classification approaches namely, decision tree induction approach (C4.5) [34], Probabilistic

Table 2
Medication to be avoided before allergy tests.

Medication	Duration
First generation antihistamines	2–3 days
Nonsedating antihistamines	7 days
Tricyclic antidepressants	7–14 days
Benzodia zepines	7–14 days
Topical corticosteroid	2–3 Weeks

approach (Naive Bayes classifier) [23], Neural Network Approach (Multi-layer Perceptron) [35], Distance based approach (k Nearest Neighbour) [36] and Support Vector Machines [37]. Classifiers are the core of a decision support system. The choice of a classifier depends on the need and purpose of the classifier for that domain of application. In allergy diagnosis, where reasoning is important, the experts and the clinicians preferred the rule base approach as it is simple and understandable. Furthermore rule based classification approaches are easier to validate when compared to neural networks and support vector machines. The classification approaches are evaluated by 10×10 -fold cross-validation using WEKA workbench with default settings [38]. Furthermore the diagnostic performance of junior clinicians was accessed in the absence of the expert. A set of test samples (test data) were given to the junior clinicians. They manually analysed the results of the IDST data and diagnosed the presence or absence of Allergic Rhinitis. The CDSS was also tested using the test data and the performance was evaluated. The complete results of the experiment and justifications are presented in the following section.

5. Experimental results

In this work four performance evaluation measures were used. The four measures namely, Precision, Sensitivity and Specificity and Accuracy differ in their evaluation focus. Precision evaluates the agreement of the class label with the positive labels predicted by the classifier. Sensitivity is used to evaluate the effectiveness of a classifier to identify positive labels whereas specificity evaluates how effectively a classifier identifies negative labels. Accuracy evaluates the overall classification efficiency of the classifier.

$$\text{Precision} = \frac{\text{Samples correctly classified as positive}}{\text{Total samples classified as positive}} \quad (4)$$

$$\text{Sensitivity} = \frac{\text{Samples correctly classified as positive}}{\text{Total positive samples in the test dataset}} \quad (5)$$

$$\text{Specificity} = \frac{\text{Samples correctly classified as negative}}{\text{Total negative samples in the test dataset}} \quad (6)$$

$$\text{Accuracy} = \frac{\text{Samples correctly classified}}{\text{Total samples classified}} \quad (7)$$

Table 3 shows the list of features selected based on correlation analysis, Information Gain (IG) and expert validation. Table 4 shows the performance of the different classifiers. The classifiers were chosen based on their approaches used for building the classification model. 10×10 -fold Cross-validation was performed to select a suitable classifier. Table 5 shows the performance of the CDSS based on feature set used and Table 6 shows the performance of the CDSS before and after oversampling the borderline minority samples.

Table 7 presents the performance of three junior clinicians in the absence of an expert. Each clinician was asked to analyse and diagnose a set of samples (test data) manually based on the IDST results. The average classification accuracy achieved by the clinicians is 58.2%.

The clinical relevance of the clinical decision support system was validated and compared with the score for allergic rhinitis (SFAR) [21]. The scoring criterion is presented in Table 8. On a scale of 16, if a sample has a score ≥ 7 , then it is classified as Allergic Rhinitis positive. The SFAR score based classification approach achieves an accuracy of 84.28%. To evaluate the performance and clinical relevance of this work, the trained CDSS is tested with the raw IDST test data. From the results presented in Table 9 it can be inferred that the CDSS without expert validation performs better than junior clinicians and the SFAR based diagnosis.

5.1. Observations and findings

Data reduction and data sampling techniques were used in the preprocessing subsystem. Feature selection is used to obtain an optimal reduced feature set. In feature selection, it can be observed that the expert has made slight alterations in the feature set to make it more relevant. For example, the feature 'Family History' is important in determining allergic disorders, but it was omitted by the mathematical reduction approaches. The initial set of features (allergens) used in this work was framed by the panel of experts at the allergy centre. Tests are continually performed at the centre for more than four decades and the allergens are also revised based on the people's present food habits and environmental conditions. For example, food items such as oats, corn, mushroom, and coco are becoming prominent food items in the daily diet for the past few years, with regard to people in Chennai and nearby rural areas. Hence the list of allergens is updated on a yearly basis.

Data sampling is used to obtain a balanced dataset. Though oversampling may slightly modify the data distribution due to addition of synthetic samples, the performance of the learning algorithm (inducer) and the model (classifier) is improved. Furthermore in medical applications the samples belonging to class of interest (positive class) is usually less which makes classification less effective. From the results, it can be inferred that the sensitivity and the overall accuracy of the classifier improves on oversampling of minority samples.

In the classification subsystem, based on comparative performance of the classification approaches, the CDSS was developed using the rule based classifier (CBA). Furthermore the rule base approach presents the knowledge model in an IF-THEN format. Hence the clinicians too prefer the rule based CDSS, as the results and justifications are more understandable than probabilistic and distance based approaches.

Figs. 2 and 3 present the contents of the rule base before and after expert validation respectively. The rule base consists of the best rules that are generated by class based association rule generation. However, it can be noticed that the antecedent of a few rules are subsets of other rules or combination of rules. For example, Rule 15 (kovaikai=0 sneeze=no $\Rightarrow$ Rhinitis=negative) is a subset of rule 7 (kovaikai=0 running nose=no sneeze=no $\Rightarrow$ Rhinitis=negative). Rule 17 (sneeze=no $\Rightarrow$ Rhinitis=negative) and rule 19 (shark=0 sneeze=no $\Rightarrow$ Rhinitis=negative) are subsets of rule 11 (shark=0 sneeze=no headache=no $\Rightarrow$ Rhinitis=negative). These rules may be removed at extra computation cost or by increasing the support and confidence thresholds, but expert validation adds more intuition and meaning to the rule base contents. It can be inferred that the modifications made by expert improves the performance of the CDSS.

Though the system framework presented in this manuscript can be adopted for the development of a DSS for a different location, the allergens (attributes) would certainly be different. Likewise, the class-based associative rule generation approach can be used for associative classification however the rules will definitely not be the same. For example, Kim et al. have developed a MAST- immunoblot assay for reliable diagnosis of allergic rhinitis [39]. They have conducted skin tests over 193 patients who visited the Yeouido St. Mary's Hospital,

Table 3
Results of feature selection.

Feature set based on correlation analysis		Feature set based on information gain		Optimal feature set based on correlation analysis, information gain and expert validation	
Age	Itching	Age	Nuckol	Aspergilus	Itching
Aspergilus	Milk(C)	Aginomoto	Old P. Dust	Beef	Kovaikai
Brinjal	Mushroom	Banana	Onion	Brinjal	Mango
Channa	Mutton	Brinjal	Paneer	Chicken	Milk(P)
Chicken	Nuckol	Channa	Prawns	Cockroach	Onion
Cockroach	Onion	Chicken	Pumpkin	Cottondust	Parthenium
Corn	Paneer	Cockroach	Red Rashes	Cough	Pollen
Cottondust	Parthenium	Coffee	Running Nose	Crab	Potato
Cough	Pepper	Corn	Gender	Curd	Potroot
Curd	Prawns	Cough	Shark	Dhal	Prawns
Dhal	Pumpkin	Dhal	Sneeze	Drumstick	Red rashes
Drumstick	Red rashes	Fish1	Spices	Egg	Running nose
Garlic	Running nose	Ginger	Swelling	F_History	Shark
Ginger	Gender	Greens	Tamarind	Fish1	Sneeze
Greens	Shark	Headache	Tea	Fish2	Swelling
Headache	Sneeze	House dust	Tomato	Ginger	Tamarind
House dust	Spices	Itching	Vazpoo/Thandu	Gram	Tea
Wheat	Swelling	Kovaikai	Wheat	Greens	Tomato
Wheezing	Tea	Maida	Wheezing	Headache	Wheezing
Yams	Tomato	Mushroom	Yams	House dust	Yams
	Rhinitis (Class)		Rhinitis (Class)		Rhinitis (Class)

Table 4
Performance evaluation of classification approaches.

	CBA	C 4.5	SVM	MLP	NB	k-NN
Sensitivity	0.940	0.894	0.934	0.899	0.931	0.961
Specificity	0.942	0.916	0.931	0.915	0.879	0.841
Precision	0.923	0.886	0.908	0.886	0.849	0.814
Accuracy(%)	94.170	90.729	93.240	90.887	90.114	89.139

Table 5
Performance evaluation of cdss based on feature selection.

	Correlation analysis	Information gain	Expert validation
Sensitivity	0.914	0.904	0.929
Specificity	0.887	0.894	0.908
Precision	0.915	0.904	0.929
Accuracy(%)	91.394	90.361	92.943

Table 6
Performance Evaluation of CDSS Based on Minority Oversampling.

	Before oversampling	After oversampling
Sensitivity	0.929	0.940
Specificity	0.908	0.942
Precision	0.929	0.923
Accuracy (%)	92.943	94.170

Table 7
Performance evaluation of junior clinicians.

	Clinician 1	Clinician 2	Clinician 3	Average
Sensitivity	0.425	0.850	0.725	0.666
Specificity	0.462	0.686	0.343	0.497
Precision	0.444	0.618	0.397	0.486
Accuracy(%)	44.400	76.790	53.410	58.200

Korea. They have used 52 allergens for skin test which are different from the set of allergens that we have used.

Hence, the generic DSS framework and the methodological approach presented in the manuscript can be used for any clinical

Table 8
Score for allergic rhinitis (SFAR).

Scoring criterion	Score	Cumulative score
Nasal blocks	1	
Running nose	1	2
Sneezing	1	3
Perennial Cough	1	4
Seasonal/Perennial	1	5
Nasal symptoms with itchy-watery eyes	2	7
House Dust trigger nasal symptoms	1	8
Pollen trigger nasal symptoms	1	9
Perceived allergic status	2	11
Previous medical diagnosis of allergy	2	13
Previous positive tests of allergy	1	14
Family history of allergy	2	16
Total score		16

Table 9
performance evaluation on raw IDST test data.

	Clinician diagnosis	SFAR based diagnosis	CDSS without expert validation	CDSS after expert validation
Sensitivity	0.666	0.874	0.856	0.883
Specificity	0.497	0.821	0.856	0.882
Precision	0.486	0.819	0.858	0.883
Accuracy (%)	58.200	84.280	85.567	88.316

DSS however the intradermal test data, allergens and rules may completely differ, so the final DSS can only be used only in Chennai and southern parts of India for diagnosis for allergic rhinitis

6. Conclusion

The prevalence of Allergic Rhinitis has significantly increased in recent years. But efficient and reliable tests for diagnosis are still lacking. Avoidance of allergens may reduce the effect of allergic symptoms. The gold standard tests used for the diagnosis of allergic diseases need an expert to validate the final outcome. In order to assist junior physicians and clinicians in the process of diagnosis of allergic diseases, this work uses a computer-aided knowledge mining

```

1.    running nose=no wheezing=no ==> Rhinitis=negative
2.    running nose=no cough=no ==> Rhinitis=negative
3.    cottondust=R sneeze=no cough=no ==> Rhinitis=negative
4.    shark=0 running nose=no cough=no ==> Rhinitis=negative
5.    pumkin=0 running nose=no sneeze=no ==> Rhinitis=negative
6.    itching=yes sneeze=no cough=no running nose=no ==> Rhinitis=negative
7.    cough=no running nose=no headache=no kovaikai=NR==> Rhinitis=negative
8.    wheeze/blocks=no running nose=no kovaikai=NR==> Rhinitis=negative
9.    itching=yes running nose=no headache=no ==> Rhinitis=negative
10.   sneeze=no running nose=no cottondust=R kovaikai=NR ==> Rhinitis=negative
11.   shark=0 kovaikai=0 running nose=no sneeze=no ==> Rhinitis=negative
12.   sneeze=yes running nose=yes housedust=R Greens=R ==> Rhinitis=positive
13.   sneeze=no running nose=no cottondust=R housedust=R ==> Rhinitis=negative
14.   sneeze=no running nose=no headache=no pollen=R ==> Rhinitis=negative
15.   sneeze=yes running nose=yes prawn=R Brinjal=R pollen=R ==> Rhinitis=positive

Fallout Rule: Rhinitis=positive

```

Fig. 3. Contents of the rule base after expert validation.

approach for extracting novel information from data. It was observed that experts and clinicians preferred a rule based approach for knowledge representation. The system is tailored to suit to the common allergens (inhalants, contactants and ingestants) prevalent in Chennai, South India. The clinical decision support system can be made even more adaptive by dynamic validation of the rule base and the rule-inference methods. The efficiency of the system can be further improved by introducing meta-heuristic data pre-processing techniques and by using ensemble classification approaches. Thus clinical decision making, among junior clinicians, can be enhanced where by allergic diseases can be diagnosed in a more precise manner. Early diagnosis and treatment can minimise the manifestation of symptoms in the patients whereby, the quality of life of the patients affected with Allergic Rhinitis is enhanced.

Acknowledgement

The authors sincerely thank Dr. L. George Moses, Allergy specialist, and the clinicians at the Good Samaritan Lab Services and Allergy Testing Centre, Kilpauk, Chennai, India, for providing the data, valuable suggestions and guidelines for this work.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.combiomed.2015.07.019>.

References

- [1] J. Bousquet, P.B. Van Cauwenberge, N. Khaltaev, N. Ait-Khaled, I. Annesi-Maesano, C. Bachert, et al., Allergic rhinitis and its impact on asthma: ARIA workshop report in collaboration with the World Health Organization, *J. Allergy Clin. Immunol.* 108 (2001) 147–334.
- [2] T. De Graaf-in't Veld, S. Koenders, I.M. Garrelds, R.G. van Wijk, The relationships between nasal hyperreactivity, quality of life, and nasal symptoms in patients with perennial allergic rhinitis, *J. Allergy Clin. Immunol.* 98 (1996) 508–513.
- [3] R. Kumar, J.K. Nagar, Raj Neelima, P. Kumar, Alka S. Kushwah, M. Meena, et al., Impact of domestic air pollution from cooking fuel on our respiratory allergies in children in India, *Asian Pac. J. Allergy Immunol.* 20 (2008) 213–222.
- [4] A. Shah, R. Pawankar, Allergic rhinitis and co-morbid asthma: perspective from India-ARIA Asia-Pacific workshop report, *Asian Pac. J. Allergy Immunol.* 27 (2009) 71–77.
- [5] S.A. Wiqar, Allergic rhinitis/rhinitis in Allergy/allergy and asthma – a clinical primer, *Indian J. Clin. Prac.* (1999) 65–70.
- [6] Ashok Shah, Ruby Pawankar, Allergic rhinitis and Co-morbid asthma: Perspective from India-ARIA Asia-Pacific Workshop Report, *Asian Pac. J. Allergy Immunol.* 27 (1) (2009) 71–77.
- [7] K. Gendo, E.B. Larson, Evidence-based diagnostic strategies for evaluating suspected allergic rhinitis, *Ann. Intern. Med.* 140 (2004) 278–289.
- [8] T.S. Jacobs, E. Forno, J.M. Brehm, E. Acosta-Pérez, Y.Y. Han, J. Blatter, J. C. Celedón, Underdiagnosis of allergic rhinitis in underserved children, *J. Allergy Clin. Immunol.* 134 (3) (2014) 737–739.
- [9] Mark S. Dykewicz, Stanley Fineman, David P. Skoner, Richard Nicklas, Rufus Lee, Joann Blessing-Moore, Diagnosis and management of rhinitis: complete guidelines of the joint task force on practice parameters in allergy, asthma and immunology, *Ann. Allergy Asthma Immunol.* 81 (1988) 478–518.
- [10] J., Oppenheimer, S., Durham, H., Nelson. Allergy diagnostic testing, URL: (http://www.worldallergy.org/professional/allergic_diseases_center/allergy_diagnostic), (2011).
- [11] J. Kellberger, H. Dressel, C. Vogelberg, W. Leupold, D. Windstetter, G. Weinmayr, J. Genuneit, C. Heumann, D. Nowak, E. von Mutius, K. Radon, Prediction of the incidence and persistence of allergic rhinitis in adolescence: a prospective cohort study, *J. Allergy Clin. Immunol.* 129 (2012) 397–402.
- [12] A. Tzuket, J. Staif, R. Levi, Y. Sharav, A computerized decision support system for patient selection in dental education, *Comput. Biol. Med.* 14 (1984) 277–284.
- [13] F.T. Tehrani, A computerized decision support system to predict the variations in the cerebral blood flow of mechanically ventilated infants, *Comput. Biol. Med.* 43 (2013) 1402–1406.
- [14] K. Sidiropoulos, D. Glotsos, S. Kostopoulos, P. Ravazoula, I. Kalatzis, D. Cavouras, J. Stonham, Real time decision support system for diagnosis of rare cancers, trained in parallel, on a graphics processing unit, *Comput. Biol. Med.* 42 (2012) 376–386.
- [15] S. Shah, A. Kusiak, Cancer gene search with data-mining and genetic algorithms, *Comput. Biol. Med.* 37 (2007) 251–261.
- [16] E. Çomak, A. Arslan, İ. Türkoğlu, A decision support system based on support vector machines for diagnosis of the heart valve diseases, *Comput. Biol. Med.* 37 (2007) 21–27.
- [17] S.J. Leslie, M. Hartswood, C. Meurig, S.P. McKee, R. Slack, R. Procter, M. A. Denvir, Clinical decision support software for management of chronic heart failure: development and evaluation, *Comput. Biol. Med.* 36 (2006) 495–506.
- [18] A. Subasi, Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines, *Comput. Biol. Med.* 42 (2012) 806–815.
- [19] M.F. El-Khatib, A diagnostic software tool for determination of complexity in respiratory pattern parameters, *Comput. Biol. Med.* 37 (2007) 1522–1527.
- [20] Anand B. Singh, Pawan Kumar, Aeroallergens in clinical practice of allergy in India. An overview, *Ann. Agric. Environ. Med.* 10 (2003) 131–136.
- [21] I. Annesi-Maesano, A. Didier, M. Klossek, I. Chanal, D. Moreau, J. Bousquet, The score for allergic rhinitis (sfar): a simple and valid assessment method in population studies, *Allergy* 57 (2002) 107–114.
- [22] Young Moon Chae, Seung Hee Ho, Chein Soo Hong, Cheol Woo Kim, Comparison of alternative knowledge model for the diagnosis of asthma, *Expert Syst. Appl.* 11 (1996) 423–429.
- [23] J. Han, M. Kamber, Data Mining, Southeast Asia Edition: Concepts and Techniques, Morgan Kaufmann, 2006, <http://store.elsevier.com/Data-Mining-Southeast-Asia-Edition/Jiawei-Han/isbn-9780080475585/>.
- [24] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *J. Artif. Intell. Res.* 16 (2002) 321–357.
- [25] H. Han, W.-Y. Wang, B.-H. Mao, Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning, *Springer* (2005) 878–887, http://link.springer.com/chapter/10.1007/11538059_91#.
- [26] J.R. Quinlan, Induction of decision trees, *Mach. Learn.* 1 (1) (1986) 81–106.
- [27] J.P. Guilford, The phi coefficient and chi square as indices of item validity, *Psychometrika* 6 (1941) 11–19.
- [28] H.O. Lancaster, E. Seneta, Chi-Square Distribution, Wiley Online Library, 1969, <http://onlinelibrary.wiley.com/doi/10.1002/0471667196.ess0238.pub2/abstract?userIsAuthenticated=false&deniedAccessCustomisedMessage=>.
- [29] B., Liu, W., Hsu, Y., Ma, Integrating classification and association rule mining, in: Proceedings of Fourth International Conference on Knowledge Discovery and Data Mining, 1998.

- [30] R., Agrawal, R., Srikant, Fast algorithms for mining association rules, in Proceedings of 20th International Conference on Very Large Data Bases, VLDB 1215, 1994, 487–499.
- [31] R., Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Ijcai*, 1995, 1137–1145.
- [32] M. De Vries, C.L. Witteman, R.W. Holland, A. Dijksterhuis, The unconscious thought effect in clinical decision making: an example in diagnosis, *Med. Decis. Mak.* 30 (2010) 578–581.
- [33] P.H. Smith, J.J. Condemi, J. Baum, R.N. Rosier, Computer-assisted measurement of skin test wheals, *J. Allergy Clin. Immunol.* 97 (1996) 208. [http://dx.doi.org/10.1016/S0091-6749\(96\)80320-4](http://dx.doi.org/10.1016/S0091-6749(96)80320-4).
- [34] J. Ross Quinlan, C 4.5: programs for machine learning, Morgan Kaufmann, San Mateo, CA (1993) c1993.
- [35] Frank Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, *Psychol. Rev.* 65 (6) (1958) 386.
- [36] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory* 13 (1) (1967) 21–27.
- [37] B.E., Boser, I.M., Guyon, V.N., Vapnik, A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 1992, 144–152.
- [38] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten, The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, 11, 1, 2009.
- [39] Y.H. Kim, B.J. Yu, W.J. Kim, J.E. Kim, G.-H. Lee, K.-A. Lee, Correlation between skin prick test and MAST-immunoblot results in patients with chronic rhinitis, *Asian Pac. J. Allergy Immunol.* 31 (2013) 20–25.

H. Khanna Nehemiah completed B.E in Computer Science and Engineering from University of Madras in 1997, M.E in Computer Science and Engineering from University of Madras in 1998 and Ph.D from Anna University, Chennai in 2007. He is currently working as an Associate Professor at Ramanujan Computing Centre, Anna University, Chennai, Tamilnadu, India. His research interests include Software Engineering, Data Mining, Database Management Systems, Bio-informatics and Medical Image Processing.

A. Kannan completed M.E in Computer Science and Engineering from Anna University Chennai in 1991 and Ph.D from Anna University, Chennai in 2000. He is currently working as Professor and Head, in the Department of Information Science and Technology, Anna University, Chennai, Tamilnadu, India. His research interests are Artificial Intelligence, Data Mining and Database Management Systems.

Jabez Christopher completed B.E. in Computer Science and Engineering from Anna University in 2009 and M.E. in Computer Science and Engineering from Anna University in 2011. He is a research scholar, pursuing PhD degree at Ramanujan Computing Centre, Anna University, Chennai, Tamilnadu, India. His research interests are Medical Data mining and Optimization.

APPENDIX I

INFORMATION GAIN and PEARSONS CORRELATION COEFFICIENT for ATTRIBUTES

INFORMATION GAIN	
0.53470442	Sneeze
0.49362142	Running Nose
0.32491318	Cough
0.30673363	Itching
0.25035326	Rashes/Swelling
0.24443921	Nasal Blocks
0.21542422	Wheezing
0.04916444	Headache
0.01605517	Brinjal
0.01086798	Housedust
0.01086597	Tea
0.00867181	Curd
0.00621485	Onion
0.00615838	Tomato
0.00585173	Ginger
0.00553569	Channa
0.00526675	Drumstick
0.00466822	Pudina
0.00465585	Pumkin
0.00435687	Fish2
0.00420019	Rice
0.00416017	Fish1
0.00396254	Aginomoto
0.00388462	Corn
0.00384992	Shark
0.00378825	Greens
0.0037262	Parthenium
0.0036742	Sex
0.003635	Coconut
0.00359676	Kovaikai
0.00316126	Cottondust
0.00316009	Pollen
0.00309483	F_History
0.00281827	Chicken
0.00278249	Coffee
0.00245655	Crab
0.00240116	Carrot
0.00235693	Beet Root
0.00227434	Oats
0.00226912	Banana
0.00221267	Cockroach
0.00219252	Radish
0.00214535	Cabbage
0.00209832	Kothavarai
0.00193525	Cauliflower
0.00191921	Spices
0.00186532	Malli
0.00184086	Mushroom
0.0018259	Maida
0.00164999	Mango
0.00155121	Prawns

0.00151171	Horlicks
0.00151171	Boost
0.00150218	Egg
0.00138844	Milk(C)
0.00136403	Gourds
0.00136212	Tamarind
0.00128659	Yams
0.00127899	Beans
0.00124351	Mutton
0.00117912	Wheat
0.00108994	Dhal
0.00088569	Garlic
0.00085959	L Finger
0.00083483	P.S Dust
0.00080978	Chillie
0.00079357	Milk(P)
0.00078253	Oil
0.00075145	Dog Fur
0.00070139	Nuckol
0.00068266	Ragi
0.00067716	Old P. Dust
0.00061981	Pepper
0.00060768	Vazpoo/Thandu
0.00043814	Potato
0.00043515	Beef
0.0003287	Potroot
0.00028816	Coco
0.00022312	Capsicum
0.00021712	Cat Dander
0.00018311	Avaraikai
0.00011837	Cucumber
0.00011375	Paneer
0.00010768	Chow-Chow
0.0000553	Gram
0.00005056	Peas
0.00001507	Aspergilus
0.0000099	Tondaikai
0.0000001	Nuts
0.0000001	Road Dust

Pearsons Correlation Coefficient	
0.80366	Sneeze
0.7844	Running Nose
0.6481	Cough
0.61974	Itching
0.55118	Rashes
0.54456	Nasal Blocks
0.53614	Wheeze/Blocks
0.26287	Headache
0.11881	Housedust
0.09689	Tea
0.09598	Curd
0.08526	Brinjal
0.08434	Tomato
0.08199	Onion
0.07137	Sex
0.07109	Pumkin
0.07102	Parthenium
0.06605	Rice
0.06559	F_History
0.06545	Pollen
0.06524	Cotondust
0.06201	Channa
0.05791	Drumstick
0.05515	Oats
0.05482	Cockroach
0.05189	Kothavarai
0.05052	Chicken
0.04956	Mushroom
0.04855	Corn
0.04735	Cabbage
0.04559	Greens
0.04544	Carrot
0.04338	Kovaikai
0.04308	Shark
0.04228	Milk(C)
0.04138	Boost
0.04138	Horlicks
0.0393	Spices
0.0391	Fish2
0.03868	Cauliflower
0.03762	Ginger
0.03703	Beet Root
0.03474	Mutton
0.03371	Beans
0.03333	Yams
0.03249	Malli
0.03074	Nuckol
0.03024	Tamarind
0.02964	Old P. Dust
0.02798	Fish1
0.02742	Vazpoo/Thandu
0.02691	Pepper
0.02515	Prawns

0.02415	Maida
0.02348	Coffee
0.02266	Dog Fur
0.02126	P.S Dust
0.02094	Potato
0.02023	Dhal
0.01985	Coco
0.01976	Garlic
0.01837	Coconut
0.01813	Egg
0.01656	Pudina
0.01601	Crab
0.01587	Potroot
0.01581	Chillie
0.01529	Capsicum
0.01503	Cat Dander
0.01499	Radish
0.01387	Gourds
0.01342	Banana
0.01272	Avaraikai
0.01221	Wheat
0.01219	Paneer
0.01143	Mango
0.00937	L Finger
0.00786	Milk(P)
0.00744	Ragi
0.00744	Peas
0.00733	Aginomoto
0.00547	Beef
0.00457	Aspergilus
0.00397	Chow-Chow
0.0037	Tondaikai
0.00185	Oil
0.00119	Cucumber
0.00116	Gram
0.00001	Nuts
0.00001	Road Dust

APPENDIX II
PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION OVER CLASS-BALANCED
IDST DATA (633 samples)

Run No.	Fold	Number of training Instances	Number of testing instances	Classifier	ACCURACY	SENSITIVITY	SPECIFICITY
1	1	569	64	CBA	92.1875	0.925926	0.918919
1	2	569	64	CBA	84.375	0.925926	0.783784
1	3	569	64	CBA	96.875	0.961538	0.973684
1	4	570	63	CBA	96.825397	0.961538	0.972973
1	5	570	63	CBA	92.063492	0.884615	0.945946
1	6	570	63	CBA	100	1	1
1	7	570	63	CBA	87.301587	0.884615	0.864865
1	8	570	63	CBA	98.412698	1	0.972973
1	9	570	63	CBA	96.825397	0.961538	0.972973
1	10	570	63	CBA	90.47619	0.961538	0.864865
2	1	569	64	CBA	92.1875	0.962963	0.891892
2	2	569	64	CBA	96.875	0.925926	1
2	3	569	64	CBA	98.4375	0.961538	1
2	4	570	63	CBA	85.714286	0.884615	0.837838
2	5	570	63	CBA	96.825397	0.961538	0.972973
2	6	570	63	CBA	95.238095	0.923077	0.972973
2	7	570	63	CBA	93.650794	0.923077	0.945946
2	8	570	63	CBA	93.650794	0.923077	0.945946
2	9	570	63	CBA	95.238095	0.961538	0.945946
2	10	570	63	CBA	98.412698	1	0.972973
3	1	569	64	CBA	96.875	1	0.945946
3	2	569	64	CBA	98.4375	1	0.972973
3	3	569	64	CBA	90.625	0.884615	0.921053
3	4	570	63	CBA	90.47619	0.884615	0.918919
3	5	570	63	CBA	98.412698	0.961538	1
3	6	570	63	CBA	90.47619	0.923077	0.891892
3	7	570	63	CBA	98.412698	0.961538	1
3	8	570	63	CBA	96.825397	1	0.945946
3	9	570	63	CBA	95.238095	0.923077	0.972973
3	10	570	63	CBA	90.47619	0.884615	0.918919
4	1	569	64	CBA	92.1875	0.962963	0.891892
4	2	569	64	CBA	92.1875	1	0.864865
4	3	569	64	CBA	95.3125	0.923077	0.973684
4	4	570	63	CBA	92.063492	0.961538	0.891892
4	5	570	63	CBA	93.650794	0.923077	0.945946
4	6	570	63	CBA	98.412698	1	0.972973
4	7	570	63	CBA	100	1	1
4	8	570	63	CBA	84.126984	0.846154	0.837838
4	9	570	63	CBA	93.650794	0.884615	0.972973
4	10	570	63	CBA	95.238095	0.923077	0.972973
5	1	569	64	CBA	87.5	0.925926	0.837838
5	2	569	64	CBA	93.75	0.888889	0.972973
5	3	569	64	CBA	93.75	0.923077	0.947368
5	4	570	63	CBA	95.238095	1	0.918919
5	5	570	63	CBA	96.825397	0.923077	1

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

5	6	570	63	CBA	96.825397	0.923077	1
5	7	570	63	CBA	96.825397	1	0.945946
5	8	570	63	CBA	90.47619	0.923077	0.891892
5	9	570	63	CBA	95.238095	0.961538	0.945946
5	10	570	63	CBA	95.238095	0.961538	0.945946
6	1	569	64	CBA	96.875	1	0.945946
6	2	569	64	CBA	89.0625	0.925926	0.864865
6	3	569	64	CBA	95.3125	0.923077	0.973684
6	4	570	63	CBA	95.238095	0.923077	0.972973
6	5	570	63	CBA	92.063492	0.884615	0.945946
6	6	570	63	CBA	95.238095	0.923077	0.972973
6	7	570	63	CBA	95.238095	0.961538	0.945946
6	8	570	63	CBA	92.063492	0.961538	0.891892
6	9	570	63	CBA	96.825397	1	0.945946
6	10	570	63	CBA	90.47619	0.884615	0.918919
7	1	569	64	CBA	96.875	1	0.947368
7	2	569	64	CBA	95.3125	0.925926	0.972973
7	3	569	64	CBA	93.75	0.925926	0.945946
7	4	570	63	CBA	96.825397	0.923077	1
7	5	570	63	CBA	96.825397	0.923077	1
7	6	570	63	CBA	90.47619	0.923077	0.891892
7	7	570	63	CBA	95.238095	0.961538	0.945946
7	8	570	63	CBA	85.714286	0.961538	0.783784
7	9	570	63	CBA	96.825397	0.923077	1
7	10	570	63	CBA	93.650794	0.923077	0.945946
8	1	569	64	CBA	89.0625	0.925926	0.864865
8	2	569	64	CBA	93.75	0.888889	0.972973
8	3	569	64	CBA	93.75	0.961538	0.921053
8	4	570	63	CBA	96.825397	1	0.945946
8	5	570	63	CBA	96.825397	0.961538	0.972973
8	6	570	63	CBA	95.238095	0.961538	0.945946
8	7	570	63	CBA	95.238095	0.923077	0.972973
8	8	570	63	CBA	92.063492	0.923077	0.918919
8	9	570	63	CBA	96.825397	0.961538	0.972973
8	10	570	63	CBA	95.238095	0.884615	1
9	1	569	64	CBA	93.75	0.961538	0.921053
9	2	569	64	CBA	93.75	0.962963	0.918919
9	3	569	64	CBA	96.875	0.962963	0.972973
9	4	570	63	CBA	98.412698	0.961538	1
9	5	570	63	CBA	100	1	1
9	6	570	63	CBA	90.47619	0.923077	0.891892
9	7	570	63	CBA	90.47619	0.884615	0.918919
9	8	570	63	CBA	95.238095	0.961538	0.945946
9	9	570	63	CBA	90.47619	0.846154	0.945946
9	10	570	63	CBA	93.650794	0.961538	0.918919
10	1	569	64	CBA	93.75	0.884615	0.973684
10	2	569	64	CBA	100	1	1
10	3	569	64	CBA	93.75	0.925926	0.945946

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

10	4	570	63	CBA	95.238095	1	0.918919
10	5	570	63	CBA	98.412698	1	0.972973
10	6	570	63	CBA	96.825397	0.961538	0.972973
10	7	570	63	CBA	90.47619	0.923077	0.891892
10	8	570	63	CBA	93.650794	1	0.891892
10	9	570	63	CBA	87.301587	0.807692	0.918919
10	10	570	63	CBA	96.825397	0.923077	1
1	1	569	64	C 4.5	90.625	0.851852	0.945946
1	2	569	64	C 4.5	85.9375	0.888889	0.837838
1	3	569	64	C 4.5	93.75	0.923077	0.947368
1	4	570	63	C 4.5	87.301587	0.884615	0.864865
1	5	570	63	C 4.5	88.888889	0.884615	0.891892
1	6	570	63	C 4.5	96.825397	0.961538	0.972973
1	7	570	63	C 4.5	87.301587	0.846154	0.891892
1	8	570	63	C 4.5	98.412698	1	0.972973
1	9	570	63	C 4.5	93.650794	0.961538	0.918919
1	10	570	63	C 4.5	85.714286	0.923077	0.810811
2	1	569	64	C 4.5	85.9375	0.925926	0.810811
2	2	569	64	C 4.5	92.1875	0.851852	0.972973
2	3	569	64	C 4.5	92.1875	0.884615	0.947368
2	4	570	63	C 4.5	90.47619	0.961538	0.864865
2	5	570	63	C 4.5	92.063492	0.846154	0.972973
2	6	570	63	C 4.5	88.888889	0.884615	0.891892
2	7	570	63	C 4.5	85.714286	0.884615	0.837838
2	8	570	63	C 4.5	88.888889	0.846154	0.918919
2	9	570	63	C 4.5	90.47619	0.961538	0.864865
2	10	570	63	C 4.5	92.063492	0.884615	0.945946
3	1	569	64	C 4.5	92.1875	0.962963	0.891892
3	2	569	64	C 4.5	98.4375	0.962963	1
3	3	569	64	C 4.5	87.5	0.846154	0.894737
3	4	570	63	C 4.5	82.539683	0.730769	0.891892
3	5	570	63	C 4.5	90.47619	0.846154	0.945946
3	6	570	63	C 4.5	85.714286	0.884615	0.837838
3	7	570	63	C 4.5	98.412698	0.961538	1
3	8	570	63	C 4.5	95.238095	1	0.918919
3	9	570	63	C 4.5	90.47619	0.884615	0.918919
3	10	570	63	C 4.5	92.063492	0.884615	0.945946
4	1	569	64	C 4.5	89.0625	0.925926	0.864865
4	2	569	64	C 4.5	90.625	0.962963	0.864865
4	3	569	64	C 4.5	98.4375	1	0.973684
4	4	570	63	C 4.5	93.650794	0.884615	0.972973
4	5	570	63	C 4.5	92.063492	0.846154	0.972973
4	6	570	63	C 4.5	96.825397	0.961538	0.972973
4	7	570	63	C 4.5	90.47619	0.884615	0.918919
4	8	570	63	C 4.5	82.539683	0.769231	0.864865
4	9	570	63	C 4.5	90.47619	0.807692	0.972973
4	10	570	63	C 4.5	92.063492	0.884615	0.945946
5	1	569	64	C 4.5	85.9375	0.851852	0.864865

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

5	2	569	64	C 4.5	85.9375	0.777778	0.918919
5	3	569	64	C 4.5	87.5	0.807692	0.921053
5	4	570	63	C 4.5	92.063492	0.961538	0.891892
5	5	570	63	C 4.5	96.825397	0.961538	0.972973
5	6	570	63	C 4.5	92.063492	0.846154	0.972973
5	7	570	63	C 4.5	96.825397	1	0.945946
5	8	570	63	C 4.5	88.888889	0.846154	0.918919
5	9	570	63	C 4.5	92.063492	0.884615	0.945946
5	10	570	63	C 4.5	95.238095	0.961538	0.945946
6	1	569	64	C 4.5	90.625	0.851852	0.945946
6	2	569	64	C 4.5	87.5	0.888889	0.864865
6	3	569	64	C 4.5	87.5	0.884615	0.868421
6	4	570	63	C 4.5	85.714286	0.807692	0.891892
6	5	570	63	C 4.5	87.301587	0.846154	0.891892
6	6	570	63	C 4.5	92.063492	0.923077	0.918919
6	7	570	63	C 4.5	92.063492	0.961538	0.891892
6	8	570	63	C 4.5	95.238095	0.961538	0.945946
6	9	570	63	C 4.5	93.650794	1	0.891892
6	10	570	63	C 4.5	88.888889	0.884615	0.891892
7	1	569	64	C 4.5	90.625	0.884615	0.921053
7	2	569	64	C 4.5	92.1875	0.851852	0.972973
7	3	569	64	C 4.5	90.625	0.851852	0.945946
7	4	570	63	C 4.5	92.063492	0.884615	0.945946
7	5	570	63	C 4.5	90.47619	0.769231	1
7	6	570	63	C 4.5	93.650794	0.961538	0.918919
7	7	570	63	C 4.5	88.888889	0.884615	0.891892
7	8	570	63	C 4.5	87.301587	1	0.783784
7	9	570	63	C 4.5	96.825397	0.923077	1
7	10	570	63	C 4.5	87.301587	0.884615	0.864865
8	1	569	64	C 4.5	79.6875	0.888889	0.72973
8	2	569	64	C 4.5	93.75	0.888889	0.972973
8	3	569	64	C 4.5	93.75	0.961538	0.921053
8	4	570	63	C 4.5	88.888889	0.884615	0.891892
8	5	570	63	C 4.5	92.063492	0.923077	0.918919
8	6	570	63	C 4.5	93.650794	0.923077	0.945946
8	7	570	63	C 4.5	93.650794	0.884615	0.972973
8	8	570	63	C 4.5	88.888889	0.884615	0.891892
8	9	570	63	C 4.5	95.238095	0.961538	0.945946
8	10	570	63	C 4.5	88.888889	0.807692	0.945946
9	1	569	64	C 4.5	93.75	0.923077	0.947368
9	2	569	64	C 4.5	90.625	0.962963	0.864865
9	3	569	64	C 4.5	92.1875	0.888889	0.945946
9	4	570	63	C 4.5	95.238095	0.923077	0.972973
9	5	570	63	C 4.5	93.650794	1	0.891892
9	6	570	63	C 4.5	90.47619	0.884615	0.918919
9	7	570	63	C 4.5	87.301587	0.846154	0.891892
9	8	570	63	C 4.5	90.47619	0.846154	0.945946
9	9	570	63	C 4.5	82.539683	0.692308	0.918919

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

9	10	570	63	C 4.5	93.650794	0.961538	0.918919
10	1	569	64	C 4.5	84.375	0.769231	0.894737
10	2	569	64	C 4.5	90.625	0.888889	0.918919
10	3	569	64	C 4.5	90.625	0.851852	0.945946
10	4	570	63	C 4.5	87.301587	0.846154	0.891892
10	5	570	63	C 4.5	90.47619	0.961538	0.864865
10	6	570	63	C 4.5	93.650794	0.923077	0.945946
10	7	570	63	C 4.5	85.714286	0.769231	0.918919
10	8	570	63	C 4.5	93.650794	1	0.891892
10	9	570	63	C 4.5	88.888889	0.884615	0.891892
10	10	570	63	C 4.5	88.888889	0.884615	0.891892
1	1	569	64	SVM	89.0625	0.851852	0.918919
1	2	569	64	SVM	90.625	0.888889	0.918919
1	3	569	64	SVM	95.3125	0.961538	0.947368
1	4	570	63	SVM	96.825397	0.961538	0.972973
1	5	570	63	SVM	90.47619	0.884615	0.918919
1	6	570	63	SVM	96.825397	1	0.945946
1	7	570	63	SVM	87.301587	0.884615	0.864865
1	8	570	63	SVM	98.412698	1	0.972973
1	9	570	63	SVM	96.825397	0.961538	0.972973
1	10	570	63	SVM	92.063492	0.961538	0.891892
2	1	569	64	SVM	85.9375	0.925926	0.810811
2	2	569	64	SVM	98.4375	0.962963	1
2	3	569	64	SVM	98.4375	0.961538	1
2	4	570	63	SVM	87.301587	0.923077	0.837838
2	5	570	63	SVM	96.825397	0.961538	0.972973
2	6	570	63	SVM	92.063492	0.923077	0.918919
2	7	570	63	SVM	90.47619	0.884615	0.918919
2	8	570	63	SVM	92.063492	0.884615	0.945946
2	9	570	63	SVM	95.238095	0.961538	0.945946
2	10	570	63	SVM	98.412698	1	0.972973
3	1	569	64	SVM	92.1875	1	0.864865
3	2	569	64	SVM	98.4375	1	0.972973
3	3	569	64	SVM	89.0625	0.923077	0.868421
3	4	570	63	SVM	90.47619	0.884615	0.918919
3	5	570	63	SVM	93.650794	0.923077	0.945946
3	6	570	63	SVM	92.063492	0.923077	0.918919
3	7	570	63	SVM	98.412698	0.961538	1
3	8	570	63	SVM	96.825397	1	0.945946
3	9	570	63	SVM	95.238095	0.923077	0.972973
3	10	570	63	SVM	85.714286	0.769231	0.918919
4	1	569	64	SVM	85.9375	0.888889	0.837838
4	2	569	64	SVM	93.75	1	0.891892
4	3	569	64	SVM	95.3125	0.923077	0.973684
4	4	570	63	SVM	98.412698	0.961538	1
4	5	570	63	SVM	87.301587	0.846154	0.891892
4	6	570	63	SVM	98.412698	1	0.972973
4	7	570	63	SVM	96.825397	1	0.945946

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

4	8	570	63	SVM	85.714286	0.884615	0.837838
4	9	570	63	SVM	95.238095	0.884615	1
4	10	570	63	SVM	93.650794	0.923077	0.945946
5	1	569	64	SVM	89.0625	0.888889	0.891892
5	2	569	64	SVM	90.625	0.888889	0.918919
5	3	569	64	SVM	90.625	0.884615	0.921053
5	4	570	63	SVM	95.238095	1	0.918919
5	5	570	63	SVM	96.825397	0.923077	1
5	6	570	63	SVM	96.825397	0.923077	1
5	7	570	63	SVM	95.238095	0.961538	0.945946
5	8	570	63	SVM	90.47619	0.961538	0.864865
5	9	570	63	SVM	93.650794	0.961538	0.918919
5	10	570	63	SVM	95.238095	0.961538	0.945946
6	1	569	64	SVM	95.3125	0.962963	0.945946
6	2	569	64	SVM	87.5	0.851852	0.891892
6	3	569	64	SVM	95.3125	0.961538	0.947368
6	4	570	63	SVM	96.825397	0.961538	0.972973
6	5	570	63	SVM	90.47619	0.884615	0.918919
6	6	570	63	SVM	93.650794	0.884615	0.972973
6	7	570	63	SVM	92.063492	0.923077	0.918919
6	8	570	63	SVM	95.238095	0.961538	0.945946
6	9	570	63	SVM	92.063492	0.961538	0.891892
6	10	570	63	SVM	93.650794	0.961538	0.918919
7	1	569	64	SVM	95.3125	0.961538	0.947368
7	2	569	64	SVM	93.75	0.888889	0.972973
7	3	569	64	SVM	93.75	0.925926	0.945946
7	4	570	63	SVM	93.650794	0.923077	0.945946
7	5	570	63	SVM	90.47619	0.846154	0.945946
7	6	570	63	SVM	88.888889	0.923077	0.864865
7	7	570	63	SVM	93.650794	0.961538	0.918919
7	8	570	63	SVM	88.888889	1	0.810811
7	9	570	63	SVM	96.825397	0.923077	1
7	10	570	63	SVM	93.650794	0.961538	0.918919
8	1	569	64	SVM	89.0625	0.925926	0.864865
8	2	569	64	SVM	92.1875	0.888889	0.945946
8	3	569	64	SVM	90.625	0.923077	0.894737
8	4	570	63	SVM	96.825397	1	0.945946
8	5	570	63	SVM	92.063492	0.884615	0.945946
8	6	570	63	SVM	95.238095	0.961538	0.945946
8	7	570	63	SVM	96.825397	0.961538	0.972973
8	8	570	63	SVM	88.888889	0.884615	0.891892
8	9	570	63	SVM	98.412698	1	0.972973
8	10	570	63	SVM	93.650794	0.923077	0.945946
9	1	569	64	SVM	96.875	1	0.947368
9	2	569	64	SVM	90.625	0.962963	0.864865
9	3	569	64	SVM	96.875	0.962963	0.972973
9	4	570	63	SVM	96.825397	0.923077	1
9	5	570	63	SVM	96.825397	1	0.945946

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

9	6	570	63	SVM	90.47619	0.923077	0.891892
9	7	570	63	SVM	90.47619	0.884615	0.918919
9	8	570	63	SVM	90.47619	0.884615	0.918919
9	9	570	63	SVM	92.063492	0.923077	0.918919
9	10	570	63	SVM	88.888889	0.846154	0.918919
10	1	569	64	SVM	96.875	0.923077	1
10	2	569	64	SVM	96.875	0.962963	0.972973
10	3	569	64	SVM	89.0625	0.925926	0.864865
10	4	570	63	SVM	95.238095	1	0.918919
10	5	570	63	SVM	95.238095	1	0.918919
10	6	570	63	SVM	93.650794	0.923077	0.945946
10	7	570	63	SVM	88.888889	0.923077	0.864865
10	8	570	63	SVM	93.650794	1	0.891892
10	9	570	63	SVM	87.301587	0.846154	0.891892
10	10	570	63	SVM	96.825397	0.923077	1
1	1	569	64	MLP	87.5	0.814815	0.918919
1	2	569	64	MLP	87.5	0.888889	0.864865
1	3	569	64	MLP	92.1875	0.961538	0.894737
1	4	570	63	MLP	95.238095	0.923077	0.972973
1	5	570	63	MLP	85.714286	0.769231	0.918919
1	6	570	63	MLP	88.888889	0.769231	0.972973
1	7	570	63	MLP	88.888889	0.923077	0.864865
1	8	570	63	MLP	95.238095	0.961538	0.945946
1	9	570	63	MLP	95.238095	0.923077	0.972973
1	10	570	63	MLP	95.238095	1	0.918919
2	1	569	64	MLP	84.375	0.888889	0.810811
2	2	569	64	MLP	92.1875	0.925926	0.918919
2	3	569	64	MLP	92.1875	0.846154	0.973684
2	4	570	63	MLP	88.888889	0.923077	0.864865
2	5	570	63	MLP	90.47619	0.769231	1
2	6	570	63	MLP	92.063492	0.923077	0.918919
2	7	570	63	MLP	92.063492	0.961538	0.891892
2	8	570	63	MLP	88.888889	0.884615	0.891892
2	9	570	63	MLP	92.063492	0.884615	0.945946
2	10	570	63	MLP	95.238095	0.923077	0.972973
3	1	569	64	MLP	90.625	0.962963	0.864865
3	2	569	64	MLP	95.3125	0.962963	0.945946
3	3	569	64	MLP	85.9375	0.846154	0.868421
3	4	570	63	MLP	90.47619	0.846154	0.945946
3	5	570	63	MLP	84.126984	0.846154	0.837838
3	6	570	63	MLP	87.301587	0.884615	0.864865
3	7	570	63	MLP	96.825397	0.961538	0.972973
3	8	570	63	MLP	98.412698	1	0.972973
3	9	570	63	MLP	87.301587	0.807692	0.918919
3	10	570	63	MLP	90.47619	0.884615	0.918919
4	1	569	64	MLP	84.375	0.851852	0.837838
4	2	569	64	MLP	93.75	0.962963	0.918919
4	3	569	64	MLP	98.4375	0.961538	1

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

4	4	570	63	MLP	93.650794	0.923077	0.945946
4	5	570	63	MLP	85.714286	0.807692	0.891892
4	6	570	63	MLP	96.825397	0.961538	0.972973
4	7	570	63	MLP	90.47619	0.961538	0.864865
4	8	570	63	MLP	82.539683	0.807692	0.837838
4	9	570	63	MLP	90.47619	0.807692	0.972973
4	10	570	63	MLP	92.063492	0.923077	0.918919
5	1	569	64	MLP	82.8125	0.851852	0.810811
5	2	569	64	MLP	87.5	0.851852	0.891892
5	3	569	64	MLP	92.1875	0.923077	0.921053
5	4	570	63	MLP	92.063492	0.961538	0.891892
5	5	570	63	MLP	95.238095	0.923077	0.972973
5	6	570	63	MLP	88.888889	0.807692	0.945946
5	7	570	63	MLP	92.063492	0.884615	0.945946
5	8	570	63	MLP	90.47619	0.923077	0.891892
5	9	570	63	MLP	88.888889	0.961538	0.837838
5	10	570	63	MLP	95.238095	1	0.918919
6	1	569	64	MLP	93.75	0.925926	0.945946
6	2	569	64	MLP	82.8125	0.740741	0.891892
6	3	569	64	MLP	89.0625	0.846154	0.921053
6	4	570	63	MLP	90.47619	0.846154	0.945946
6	5	570	63	MLP	88.888889	0.807692	0.945946
6	6	570	63	MLP	93.650794	0.923077	0.945946
6	7	570	63	MLP	93.650794	0.961538	0.918919
6	8	570	63	MLP	92.063492	1	0.864865
6	9	570	63	MLP	95.238095	0.961538	0.945946
6	10	570	63	MLP	88.888889	0.807692	0.945946
7	1	569	64	MLP	92.1875	0.923077	0.921053
7	2	569	64	MLP	93.75	0.888889	0.972973
7	3	569	64	MLP	93.75	0.962963	0.918919
7	4	570	63	MLP	95.238095	0.923077	0.972973
7	5	570	63	MLP	90.47619	0.846154	0.945946
7	6	570	63	MLP	88.888889	0.884615	0.891892
7	7	570	63	MLP	87.301587	0.884615	0.864865
7	8	570	63	MLP	87.301587	0.961538	0.810811
7	9	570	63	MLP	98.412698	0.961538	1
7	10	570	63	MLP	88.888889	0.884615	0.891892
8	1	569	64	MLP	85.9375	0.925926	0.810811
8	2	569	64	MLP	89.0625	0.814815	0.945946
8	3	569	64	MLP	92.1875	0.923077	0.921053
8	4	570	63	MLP	92.063492	0.961538	0.891892
8	5	570	63	MLP	90.47619	0.923077	0.891892
8	6	570	63	MLP	93.650794	0.961538	0.918919
8	7	570	63	MLP	92.063492	0.846154	0.972973
8	8	570	63	MLP	88.888889	0.884615	0.891892
8	9	570	63	MLP	93.650794	0.923077	0.945946
8	10	570	63	MLP	96.825397	0.961538	0.972973
9	1	569	64	MLP	93.75	1	0.894737

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

9	2	569	64	MLP	84.375	0.851852	0.837838
9	3	569	64	MLP	95.3125	1	0.918919
9	4	570	63	MLP	93.650794	0.884615	0.972973
9	5	570	63	MLP	96.825397	0.961538	0.972973
9	6	570	63	MLP	84.126984	0.846154	0.837838
9	7	570	63	MLP	92.063492	0.884615	0.945946
9	8	570	63	MLP	88.888889	0.846154	0.918919
9	9	570	63	MLP	90.47619	0.884615	0.918919
9	10	570	63	MLP	85.714286	0.807692	0.891892
10	1	569	64	MLP	92.1875	0.846154	0.973684
10	2	569	64	MLP	96.875	1	0.945946
10	3	569	64	MLP	87.5	0.888889	0.864865
10	4	570	63	MLP	95.238095	1	0.918919
10	5	570	63	MLP	90.47619	0.923077	0.891892
10	6	570	63	MLP	90.47619	0.923077	0.891892
10	7	570	63	MLP	84.126984	0.846154	0.837838
10	8	570	63	MLP	92.063492	0.961538	0.891892
10	9	570	63	MLP	82.539683	0.730769	0.891892
10	10	570	63	MLP	92.063492	0.846154	0.972973
1	1	569	64	Bayes	90.625	0.888889	0.918919
1	2	569	64	Bayes	81.25	0.888889	0.756757
1	3	569	64	Bayes	92.1875	0.923077	0.921053
1	4	570	63	Bayes	87.301587	0.846154	0.891892
1	5	570	63	Bayes	92.063492	0.884615	0.945946
1	6	570	63	Bayes	90.47619	1	0.837838
1	7	570	63	Bayes	87.301587	0.884615	0.864865
1	8	570	63	Bayes	93.650794	1	0.891892
1	9	570	63	Bayes	95.238095	0.961538	0.945946
1	10	570	63	Bayes	90.47619	0.961538	0.864865
2	1	569	64	Bayes	81.25	0.888889	0.756757
2	2	569	64	Bayes	96.875	1	0.945946
2	3	569	64	Bayes	90.625	0.923077	0.894737
2	4	570	63	Bayes	87.301587	0.961538	0.810811
2	5	570	63	Bayes	95.238095	0.923077	0.972973
2	6	570	63	Bayes	90.47619	0.884615	0.918919
2	7	570	63	Bayes	87.301587	0.923077	0.837838
2	8	570	63	Bayes	84.126984	0.884615	0.810811
2	9	570	63	Bayes	90.47619	0.961538	0.864865
2	10	570	63	Bayes	96.825397	0.961538	0.972973
3	1	569	64	Bayes	89.0625	0.962963	0.837838
3	2	569	64	Bayes	93.75	0.888889	0.972973
3	3	569	64	Bayes	85.9375	0.884615	0.842105
3	4	570	63	Bayes	87.301587	0.961538	0.810811
3	5	570	63	Bayes	93.650794	0.961538	0.918919
3	6	570	63	Bayes	88.888889	0.961538	0.837838
3	7	570	63	Bayes	95.238095	0.961538	0.945946
3	8	570	63	Bayes	88.888889	0.923077	0.864865
3	9	570	63	Bayes	93.650794	0.961538	0.918919

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

3	10	570	63	Bayes	85.714286	0.846154	0.864865
4	1	569	64	Bayes	87.5	0.962963	0.810811
4	2	569	64	Bayes	82.8125	0.925926	0.756757
4	3	569	64	Bayes	93.75	0.923077	0.947368
4	4	570	63	Bayes	90.47619	0.923077	0.891892
4	5	570	63	Bayes	85.714286	0.884615	0.837838
4	6	570	63	Bayes	88.888889	0.884615	0.891892
4	7	570	63	Bayes	96.825397	1	0.945946
4	8	570	63	Bayes	88.888889	0.961538	0.837838
4	9	570	63	Bayes	90.47619	0.923077	0.891892
4	10	570	63	Bayes	95.238095	0.961538	0.945946
5	1	569	64	Bayes	84.375	0.851852	0.837838
5	2	569	64	Bayes	92.1875	0.962963	0.891892
5	3	569	64	Bayes	92.1875	0.923077	0.921053
5	4	570	63	Bayes	87.301587	1	0.783784
5	5	570	63	Bayes	92.063492	0.884615	0.945946
5	6	570	63	Bayes	95.238095	0.961538	0.945946
5	7	570	63	Bayes	93.650794	0.923077	0.945946
5	8	570	63	Bayes	90.47619	0.923077	0.891892
5	9	570	63	Bayes	85.714286	0.884615	0.837838
5	10	570	63	Bayes	90.47619	1	0.837838
6	1	569	64	Bayes	96.875	1	0.945946
6	2	569	64	Bayes	85.9375	0.851852	0.864865
6	3	569	64	Bayes	87.5	0.961538	0.815789
6	4	570	63	Bayes	90.47619	0.923077	0.891892
6	5	570	63	Bayes	87.301587	0.884615	0.864865
6	6	570	63	Bayes	90.47619	0.884615	0.918919
6	7	570	63	Bayes	93.650794	0.923077	0.945946
6	8	570	63	Bayes	87.301587	1	0.783784
6	9	570	63	Bayes	93.650794	1	0.891892
6	10	570	63	Bayes	85.714286	0.884615	0.837838
7	1	569	64	Bayes	90.625	0.923077	0.894737
7	2	569	64	Bayes	90.625	0.925926	0.891892
7	3	569	64	Bayes	92.1875	0.888889	0.945946
7	4	570	63	Bayes	90.47619	0.884615	0.918919
7	5	570	63	Bayes	90.47619	0.923077	0.891892
7	6	570	63	Bayes	88.888889	0.923077	0.864865
7	7	570	63	Bayes	93.650794	0.961538	0.918919
7	8	570	63	Bayes	85.714286	1	0.756757
7	9	570	63	Bayes	90.47619	0.961538	0.864865
7	10	570	63	Bayes	92.063492	0.961538	0.891892
8	1	569	64	Bayes	81.25	0.851852	0.783784
8	2	569	64	Bayes	90.625	0.888889	0.918919
8	3	569	64	Bayes	85.9375	0.884615	0.842105
8	4	570	63	Bayes	88.888889	1	0.810811
8	5	570	63	Bayes	93.650794	0.961538	0.918919
8	6	570	63	Bayes	92.063492	1	0.864865
8	7	570	63	Bayes	96.825397	0.961538	0.972973

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

8	8	570	63	Bayes	87.301587	0.884615	0.864865
8	9	570	63	Bayes	95.238095	0.961538	0.945946
8	10	570	63	Bayes	88.888889	0.923077	0.864865
9	1	569	64	Bayes	92.1875	1	0.868421
9	2	569	64	Bayes	87.5	0.962963	0.810811
9	3	569	64	Bayes	92.1875	0.962963	0.891892
9	4	570	63	Bayes	96.825397	0.961538	0.972973
9	5	570	63	Bayes	90.47619	0.923077	0.891892
9	6	570	63	Bayes	87.301587	0.923077	0.837838
9	7	570	63	Bayes	90.47619	0.961538	0.864865
9	8	570	63	Bayes	88.888889	0.884615	0.891892
9	9	570	63	Bayes	90.47619	0.884615	0.918919
9	10	570	63	Bayes	82.539683	0.884615	0.783784
10	1	569	64	Bayes	93.75	0.923077	0.947368
10	2	569	64	Bayes	93.75	0.962963	0.918919
10	3	569	64	Bayes	85.9375	0.888889	0.837838
10	4	570	63	Bayes	92.063492	1	0.864865
10	5	570	63	Bayes	92.063492	1	0.864865
10	6	570	63	Bayes	92.063492	0.961538	0.891892
10	7	570	63	Bayes	87.301587	0.884615	0.864865
10	8	570	63	Bayes	85.714286	0.884615	0.837838
10	9	570	63	Bayes	87.301587	0.884615	0.864865
10	10	570	63	Bayes	90.47619	0.884615	0.918919
1	1	569	64	kNN	84.375	0.962963	0.756757
1	2	569	64	kNN	82.8125	0.888889	0.783784
1	3	569	64	kNN	87.5	0.961538	0.815789
1	4	570	63	kNN	90.47619	0.923077	0.891892
1	5	570	63	kNN	92.063492	0.961538	0.891892
1	6	570	63	kNN	92.063492	1	0.864865
1	7	570	63	kNN	87.301587	0.923077	0.837838
1	8	570	63	kNN	90.47619	1	0.837838
1	9	570	63	kNN	92.063492	0.961538	0.891892
1	10	570	63	kNN	92.063492	1	0.864865
2	1	569	64	kNN	79.6875	0.925926	0.702703
2	2	569	64	kNN	95.3125	1	0.918919
2	3	569	64	kNN	90.625	1	0.842105
2	4	570	63	kNN	84.126984	0.961538	0.756757
2	5	570	63	kNN	95.238095	1	0.918919
2	6	570	63	kNN	93.650794	0.961538	0.918919
2	7	570	63	kNN	87.301587	1	0.783784
2	8	570	63	kNN	87.301587	0.884615	0.864865
2	9	570	63	kNN	84.126984	0.923077	0.783784
2	10	570	63	kNN	95.238095	0.961538	0.945946
3	1	569	64	kNN	87.5	0.925926	0.837838
3	2	569	64	kNN	92.1875	1	0.864865
3	3	569	64	kNN	78.125	0.923077	0.684211
3	4	570	63	kNN	90.47619	1	0.837838
3	5	570	63	kNN	87.301587	0.961538	0.810811

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

3	6	570	63	kNN	88.888889	1	0.810811
3	7	570	63	kNN	96.825397	1	0.945946
3	8	570	63	kNN	98.412698	1	0.972973
3	9	570	63	kNN	88.888889	0.961538	0.837838
3	10	570	63	kNN	84.126984	0.846154	0.837838
4	1	569	64	kNN	84.375	1	0.72973
4	2	569	64	kNN	85.9375	0.962963	0.783784
4	3	569	64	kNN	90.625	1	0.842105
4	4	570	63	kNN	88.888889	0.961538	0.837838
4	5	570	63	kNN	84.126984	0.923077	0.783784
4	6	570	63	kNN	87.301587	0.884615	0.864865
4	7	570	63	kNN	92.063492	1	0.864865
4	8	570	63	kNN	85.714286	0.884615	0.837838
4	9	570	63	kNN	92.063492	0.961538	0.891892
4	10	570	63	kNN	93.650794	0.961538	0.918919
5	1	569	64	kNN	84.375	0.962963	0.756757
5	2	569	64	kNN	84.375	0.925926	0.783784
5	3	569	64	kNN	90.625	1	0.842105
5	4	570	63	kNN	87.301587	1	0.783784
5	5	570	63	kNN	90.47619	0.923077	0.891892
5	6	570	63	kNN	93.650794	1	0.891892
5	7	570	63	kNN	96.825397	1	0.945946
5	8	570	63	kNN	84.126984	0.923077	0.783784
5	9	570	63	kNN	85.714286	0.923077	0.810811
5	10	570	63	kNN	90.47619	0.961538	0.864865
6	1	569	64	kNN	93.75	1	0.891892
6	2	569	64	kNN	85.9375	0.925926	0.810811
6	3	569	64	kNN	84.375	0.884615	0.815789
6	4	570	63	kNN	93.650794	1	0.891892
6	5	570	63	kNN	87.301587	0.923077	0.837838
6	6	570	63	kNN	92.063492	0.961538	0.891892
6	7	570	63	kNN	93.650794	1	0.891892
6	8	570	63	kNN	87.301587	0.961538	0.810811
6	9	570	63	kNN	90.47619	1	0.837838
6	10	570	63	kNN	92.063492	1	0.864865
7	1	569	64	kNN	87.5	0.961538	0.815789
7	2	569	64	kNN	85.9375	0.962963	0.783784
7	3	569	64	kNN	93.75	0.962963	0.918919
7	4	570	63	kNN	90.47619	0.884615	0.918919
7	5	570	63	kNN	87.301587	1	0.783784
7	6	570	63	kNN	88.888889	0.961538	0.837838
7	7	570	63	kNN	85.714286	0.961538	0.783784
7	8	570	63	kNN	84.126984	1	0.72973
7	9	570	63	kNN	90.47619	0.961538	0.864865
7	10	570	63	kNN	93.650794	0.961538	0.918919
8	1	569	64	kNN	89.0625	0.962963	0.837838
8	2	569	64	kNN	87.5	0.962963	0.810811
8	3	569	64	kNN	87.5	1	0.789474

PERFORMANCE EVALUATION USING 10x10 FOLD CROSS VALIDATION

8	4	570	63	kNN	88.888889	1	0.810811
8	5	570	63	kNN	92.063492	0.961538	0.891892
8	6	570	63	kNN	95.238095	1	0.918919
8	7	570	63	kNN	95.238095	0.961538	0.945946
8	8	570	63	kNN	82.539683	0.884615	0.783784
8	9	570	63	kNN	85.714286	1	0.756757
8	10	570	63	kNN	90.47619	0.923077	0.891892
9	1	569	64	kNN	87.5	0.923077	0.842105
9	2	569	64	kNN	84.375	0.925926	0.783784
9	3	569	64	kNN	90.625	1	0.837838
9	4	570	63	kNN	96.825397	1	0.945946
9	5	570	63	kNN	88.888889	0.961538	0.837838
9	6	570	63	kNN	88.888889	1	0.810811
9	7	570	63	kNN	88.888889	0.923077	0.864865
9	8	570	63	kNN	90.47619	1	0.837838
9	9	570	63	kNN	88.888889	0.961538	0.837838
9	10	570	63	kNN	88.888889	0.961538	0.837838
10	1	569	64	kNN	87.5	0.961538	0.815789
10	2	569	64	kNN	87.5	1	0.783784
10	3	569	64	kNN	89.0625	0.962963	0.837838
10	4	570	63	kNN	92.063492	1	0.864865
10	5	570	63	kNN	92.063492	1	0.864865
10	6	570	63	kNN	87.301587	0.961538	0.810811
10	7	570	63	kNN	82.539683	0.923077	0.756757
10	8	570	63	kNN	93.650794	0.961538	0.918919
10	9	570	63	kNN	80.952381	0.846154	0.783784
10	10	570	63	kNN	95.238095	1	0.918919