

Acknowledgements

I thank my supervisor *Dr. Ganesh Bagler* for making this thesis a reality. I am indebted to him for his constant encouragement and commitment to perfection in research. This thesis is a result of many collaborations. I want to thank each and everyone who contributed towards this work, specifically *Anupam Jain* and *Rudraksh Tuwani*.

I thank the academic and administrative staff at the Indian Institute of Technology Jodhpur for helping me during my period of research at the institute. I am thankful to all the faculty members who were part of my PhD committee, *Dr. Vidya Saroeswaran*, *Dr. Kirankumar Hiremath* and *Dr. Sushmita Paul*, for their constructive criticism and support. I extend my deep love and gratitude towards my fellow lab members and friends-*Rahul Badhwar*, *Rakesh Kanji*, *Megha Singh*, *Ridhi Aggarwal*, *Bhuvnesh Rathore*, *Shalini Singh*, and all my friends at IIT Jodhpur for their help and support throughout the entire program.

I extend my gratitude to everyone at IIIT Delhi for the help I received during my stay. I am indebted to my *Parents* and my brother *Rakesh* for providing comfort during the most troubled times. Last but not least, my husband *Renjith* for taking part in this difficult journey with me.

Rakhi N K
PhD. Student

List of Figures

Figures	Title	Page
3.1	Overview of Indian Cuisine Data	27
3.2	Compilation of Flavor Compounds Data	28
3.3	Distribution of Flavor Profiles of ten major Ingredient Categories	28
3.4	Statistics of Indian cuisine	29
4.1	Illustration of Procedure used for computation of average food pairing of a cuisine	32
4.2	Strong negative food pairing in Indian cuisine and constituent sub-cuisines	33
4.3	Negative food pairing at ingredient level and investigation of food pairing with recipe-level statistics	35
4.4	Spices are the critical contributors to negative food pairing in Indian cuisine	35
5.1	Recipe size distributions	38
5.2	Frequency-Rank distributions	39
5.3	ΔN_s and its statistical significance	40
5.4	Strong negative food pairing in Indian cuisine and constituent sub-cuisines	41
5.5	Cumulative probability distribution of N_s^R values for regional cuisines and their random controls	43
5.6	Co-occurrence of ingredients with increasing extent of flavor profile overlap	44
5.7	Contribution of ingredients (χ_i) towards flavor pairing	46
5.8	Contribution of individual categories (ΔN_s^{cat}) towards food pairing bias and its statistical significance.	47
5.9	Variation in category contribution and its statistical significance	48
6.1	FlavorDB Overview	50
6.2	Schematic of FlavorDB user interface highlighting features for searching and graphical navigation of data	54
7.1	Workflow implemented for data-driven analysis of biomedical literature associating culinary spices and herbs to diseases.	58
7.2	Hierarchical structure of MeSH disease headers	60
7.3	Architecture of the Convolutional Neural Network	62
7.4	Statistics of spice-disease associations: Historical trend in biomedical literature reporting spice- disease associations	63
7.5	Statistics of positive and negative disease associations for the top 50 spices with most number of associations	64
7.6	Top diseases (Third level of MeSH hierarchy) ranked according to their total number of positive associations	65
7.7	Top diseases (Third level of MeSH hierarchy) ranked according to their total number of negative associations	65
8.1	Disease categories (First level of MeSH hierarchy) ranked according to the number of positive associations with spices	70
8.2	Disease categories (First level of MeSH hierarchy) ranked according to the number of negative associations with spices	70
8.3	Spices ranked according to their 'relative benevolence score' highlighting their broad-spectrum benevolence	72
9.1	Overview of SpiceRx	78
9.2	The user interface of SpiceRx	80
9.3	Searching for drug-like phytochemicals in SpiceRx fulfilling Lipinski's rule	82

List of Tables

<i>Table</i>	<i>Title</i>	<i>page</i>
3.1	Number of ingredients in each category. Ten major categories with five or more ingredients are highlighted.	26
5.1	Statistics of regional cuisines	37
5.2	List of ingredient categories and corresponding ingredient counts as found in all sub-cuisines.	38
7.1	The class-wise performance metrics for the best CNN model, implementing word, position, part of speech and chunk embedding features, used for spice-disease relationship extraction	62

List of Symbols

Symbol	Description
R	Recipe
S	Size of the recipe
N_R	Number of recipes
N_I	Number of Ingredients
F_i	Flavor profile of ingredient i
N_S^R	Flavor sharing in a recipe
$\overline{N_S}$	Flavor sharing of the cuisine
ΔN_S	Average flavor sharing index
$f(N)$	Fraction of ingredient pairs frequency
P	Probability
χ^i	Magnitude Of Average Food Pairing
\widehat{D}_s	Number of disease categories with which spice s has therapeutic association with
\widehat{d}_s^i	Number of disease subcategories in the i th disease category with which the spice s is associated
d^i	Total number of disease sub-categories in the i th disease category
$\widehat{\alpha}_s^k$	Number of disease subcategories in the k th disease sub-category with which the spice s is associated.
Ω_s^+	Benevolence Spectrum Score
$\Delta \Omega_s$	Relative Benevolence Score
$\Delta \Omega_s^i$	Therapeutic tradeoff score
Ω_s^j	Category-specific (benevolence and adverse) spectrum score
σ	Standard deviation

List of Abbreviations

Abbreviation *Full form*

API	Application Program Interphase
CAS	Chemical Abstract Service
CNN	Convolutional Neural Network
CTD	Comparative Toxicogenomics Database
FEMA	Flavor Extract Manufactures Association
FN	False Negative
FP	False Positive
FPI	Food Plants International
IC	Indian Cuisine
IUPAC	International Union for Pure and Applied Chemistry
MESH	Medical Subject Headings
NER	Named Entity Recognition
NLP	Natural Language Processing
PFAF	Plants for a Future
PMID	PubMed ID
POS	Parts of Speech
SDF	Spatial Data File
SMILES	Simple Molecular Input Line Entry System
SVM	Support Vector Machines
TF-IDF	Term Frequency-Inverse Document Frequency
TN	True Negative
TP	True Positive

