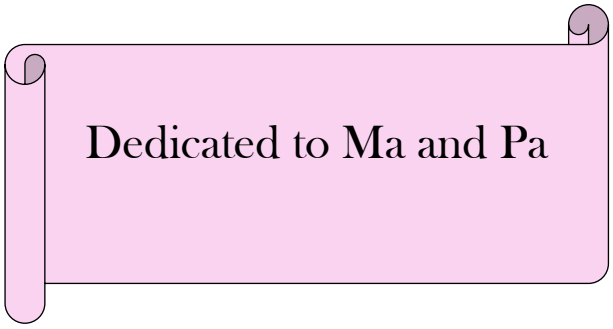
I know what it is to be in need, and I know what it is to have plenty. I have learned the secret of being content in any and every situation, whether well fed or hungry, whether living in plenty or in want.
I can do everything through him who gives me strength.



Declaration by the candidate

The candidate certifies that the thesis entitled **"Computational gastronomic study on flavour pairing behaviour in food recipes"** submitted to the **School of Engineering, Tezpur University** in partial fulfilment for the award of the degree of **Doctor of Philosophy** in the **Department of Food Engineering and Technology** is a record of research work carried out by me under the supervision of **Prof. Manuj Kumar Hazarika**.

All assistance received from various sources have been appropriately acknowledged. No part of this thesis has been submitted elsewhere for the award of any degree.

Date:

Place:

L.V Makinei
Reg. No TZ200640, Roll No. FPP16103
Department of Food Engineering and Technology
School of Engineering, Tezpur University
Napaam 784028, India

तेजपुर विश्वविद्यालय
(केंद्रीय विश्वविद्यालय)
नपाम, तेजपुर, असम, ७८४०२८

TEZPUR UNIVERSITY
(A Central University)
Napaam, Tezpur, Assam
PIN: 784028, India

कुलाध्यक्ष का सर्वोत्तम विश्वविद्यालय पुरस्कार, २०१६
Visitor's Best University Award 2016

Dr. M K Hazarika,
Professor,
Department of Food Engineering & Technology.

Email: mkhazarika@tezu.ernet.in
Ph:- +91 3712 275706

Dated: …/…../……..

Certificate from the Supervisor

This is to certify that the thesis entitled **"Computational gastronomic study on flavour pairing behaviour in food recipes"** submitted to the **School of Engineering, Tezpur University** in partial fulfillment for the award of the degree of **Doctor of Philosophy** in the **Department of Food Engineering and Technology** is a record of research work carried out by **Ms. L.V. Makinei** under my supervision and guidance.

All assistance received by the candidate from various sources has been appropriately acknowledged. No part of this thesis has been submitted elsewhere for award of any other degree.

Date:
Place:

Prof. Manuj Kumar Hazarika
Department of food Engineering and Technology
School of Engineering, Tezpur University
Napaam 784028, Assam (India)

Acknowledgements

First and foremost, I thank God for His grace and mercy that have led me to this very moment.

With deep gratitude, I would like to express my sincere thanks to my supervisor Prof. Manuj Kumar Hazarika. It is because of his help, time and academic guidance that I have been able to complete my research work. It has been a great experience to work under him.

I would like to thank my Doctoral Committee members, Prof. Manabendra Mandal and Prof. Poonam Mishra for their encouragement, and valuable suggestions at various stages of my research. I also thank the Doctoral research committee members for their insightful comment and guidance throughout my research work. I also acknowledge the entire faculties, technical officer, technician, office staff of Department of Food Engineering and Technology for their constant help and support.

I would also like to acknowledge the financial support from University Grants Commission (UGC NET-JRF) for supporting me throughout my research.

I express my sincere thanks to my lab junior Ms. Shagufta Rizwana for her constant help and support throughout my research work. I also thank my lab mates Swapnil Gautam, Tridisha Bordoloi, Sonam Kumari and my Juniors for their motivation and encouragement. Also, a special shout out to my friend Dr. Chiging Yamang for being my confidante.

I would like to express my gratitude to my Parents for their constant emotional support and for believing in me. My siblings, Lamak, Noel, Samzeila and Apollos for supporting me always. My relatives, uncles, aunts, cousins, and friends for all their prayers, support, interest, motivation and encouragement to persevere through the difficult times of my research. A special thanks to my cousin brothers Mr. Rejoyson Thangal and Dr. Moses who helped me in their own special ways.

Finally, I express my thanks to all whose names have not been mentioned individually but have helped me directly or indirectly in this work.

Date: L.V. Makinei

Place:

LIST OF TABLES

Table No.	Title	Page No.
2.1	Summary from major reported studies based on chronological development in computational gastronomy and customized food recommendation	29-30
3.1	Format for tabulation of recipe data	32
3.2	Format for tabulation of flavor compound data	34
3.3	List of software used	35
4.1	Regional cuisine statistics of recipe data	62
4.2	Number of ingredients in each category	62
4.3	Statistics of different datasets	63
4.4	List of additional ingredients from Northeast	63
4.5	Statistics of food pairing behaviour	73
4.6	Similarity percentage of recipes across regional cuisine	88
4.7	Number of similar recipes from the Northeast and other Indian regional cuisines	90
4.8	Number of similar recipes of the Northeast regional recipes compared to western and east Asian	90
4.9	Average shared compounds of ingredients	95
4.10	List of spices for new recipe	96
4.11	Top 10 most prevalent ingredients in khichdi recipe	98
4.12	List of authentic ingredients to be added to the set of prevalent ingredients of khichdi recipe for new recipe generation	99
4.13	Three best paired Ingredient of the regional cuisines	102
4.14	Pairing recommendation to complete the set of three best paired ingredients	103
4.15	Alternative ingredients for pork in pork with bamboo shoot curry Cosine similarity table of the recommended recipe set	106
4.16	Alternative ingredients for bamboo shoot in khorisar lagot gahori mangxo	106
4.17	Alternative ingredients for soybean in dawlrep bai (mizo stew)	107
4.18	Alternative ingredients for potato in ironba (manipur chutney)	107

LIST OF FIGURES

Figure No.	Title	Page No.
1.1	Graphical representation of the thesis structure	8
3.1	Steps for data pre-processing	33
3.2	Python libraries used for pre-processing of recipe data	34
3.3	Cytoscape 3.8.3 desktop	36-37
3.4	Cytoscape 3.8.3 import network data files	38
3.5	Cytoscape 3.8.3 layout styles	39
3.6	Cytoscape 3.8.3 export images	39
3.7	Cytoscape 3.8.3 visual styles	40
3.8	Cytoscape 3.8.3 mapping styles	40
3.9	Illustration of data file compilation	41
3.10	Average shared compound calculation for real and random cuisine	42
3.11	Random control generation	43
3.12	Python libraries used for creating random controls	44
3.13	Random recipe generation	45
3.14	Python libraries used for cosine similarity analysis	46
3.15	Python code for calculating TF-IDF	46
3.16	Python libraries used for t-SNE analysis	46
3.17	Python code for t-SNE clustering	47
3.18	Python code for plotting t-SNE in flavour and ingredient space	47
3.19	Illustration for calculating average flavour sharing $N_s(R)$ on recipe level	49
3.20	Illustration for calculating flavour sharing index $\overline{N}_s$ on cuisine level	50
3.21	Illustration of analysis based on flavour similarity	54
3.22	Flow chart of the implementation process of the data-driven models for recipe completion	56
3.23	General concept for the alternative ingredient selection based on the similarity and compatibility of ingredients	57
4.1(a)	Regional cuisine recipe size distribution	65
4.1(b)	Regional cuisine frequency rank distribution	65

4.1(c)	Regional cuisine complementary cumulative degree distribution plot with power-law fitting $k^{-0.77}$	65
4.2	Flavour pyramids for the Northeast regional cuisines	69
4.3	Frequency rank distribution with its corresponding random models	71
4.4	Statistical significance of ΔNs, which indicates the extent of bias in food pairings of regional cuisines with their random models	74
4.5 (a)	Ingredient contribution Assam	75
4.5 (b)	Ingredient contribution Arunachal	75
4.5 (c)	Ingredient contribution Manipur	76
4.5 (d)	Ingredient contribution Meghalaya	75
4.5 (e)	Ingredient contribution Mizoram	77
4.5 (f)	Ingredient contribution Nagaland	75
4.5 (g)	Ingredient contribution Sikkim	78
4.5 (h)	Ingredient contribution Tripura	75
4.6	t-SNE clustering of Northeast regional cuisine in ingredient space and flavour space	82
4.7	t-SNE clustering of Northeast regional cuisine and Indian cuisine in ingredient space and flavour space	82
4.8	t-SNE clustering of Northeast regional cuisine and other countries in ingredient space and flavour space	83
4.9	The ingredient-compound bipartite network of an authentic ingredient triplet	84
4.10	The flavour graph backbone of 126 ingredients of Northeast cuisine. A node represents each of the 126 ingredients, and an edge represents a shared flavour compound. sizes of the nodes are scaled based on the frequency of use of ingredients, while edges are scaled based on the amount of flavour compounds shared.	86
4.11(a)	Size of each node represents the prevalence of the ingredients; the width of edges represents the number of shared flavour compounds between the ingredients and the colour of the node represents the food category	87
4.11(b)	Size of each node represents the number of flavour compounds of the ingredients; the width of edges represents the shared flavour compounds between the ingredients and the colour of the node represents the food category	87
4.12 (a)	Flavour network plot of recipe chital fish fry and ari fish fry with a difference of one ingredient in between them	92

4.12 (b)	Flavour network plot of recipe black gram with ou tenga and black gram curry with a difference of one ingredient in between them	92
4.12 (c)	Flavour network plot of caramel pudding (Northeast recipe) and western_92 (western recipe)	93
4.12 (d)	Flavour network plot of arsa pok (Northeast recipe) with east Asian17 (east Asian recipe)	93
4.13	Nine-point hedonic score of the new recipe sensory analysis	97
4.14	Cosine similarity of new recipe and pre-existing recipe	98
4.15	Flavour network of Khichdi recipe; the size of the node represents the prevalence of the ingredients and the size/thickness of the link between the ingredients represents the number of shared flavour compounds, colour of the nodes represents the ingredients categories	99
4.16	Flavour cosine similarity of recipes across regional cuisines which is the result of flavour network	100
4.17 (a-h)	Flavour cosine similarity of recipes across regional cuisines with top seven prevalent ingredients selected from the flavour network, AU denotes authentic ingredients	101-102
4.18 (a)	Cosine similarity with original recipe pork replaced with mutton	108
4.18 (b)	Cosine similarity with original recipe soybean replaced with potato	108

List of Notations and Abbreviation

χ_i	The contribution of i^{th} ingredient in a cuisine
k	Number of recipes
s	The number of ingredients in a recipe
t-SNE	t-Distributed Stochastic Neighbour Embedding
C	Any given cuisine
CSV	Comma-separated value, format for tabulated data
NMF	Non-negative matrix factorization
R	Represent a recipe
RLS	Recursive least squares
TF-IDF	Term frequency inverse document frequency
TSV	Tab-separated value, format for tabulated data
F_i	Set of flavour compounds found in the ingredient i
F_j	Set of flavour compounds found in the ingredient j
N_c	Number of recipes in the cuisine C
$\bar{N}_s$	The average number of shared flavouring compounds at cuisine level
$N_s(R)$	The average number of shared flavouring compounds at recipe level
$\bar{N}_s^{Real}$	$\bar{N}_s$ values of a real cuisine
$\bar{N}_s^{Rand}$	$\bar{N}_s$ values of a random cuisine
ΔN_S	The difference between the average number of shared flavouring compounds of the real cuisine and a random cuisine
$\bar{N}_s(C)$	The degree of the flavour pairing of a given cuisine C
$\bar{N}_s(C^i)$	The degree of the flavour pairing of the cuisine C after the exclusion of ingredient of concern i
p_i^C	Authenticity of an ingredient ***i*** in cuisine C
p_{ij}^C	Authenticity of an ingredient pair ***i,j*** in cuisine C
p_{ijk}^C	Authenticity of an ingredient triplet ***i,j,k*** in cuisine C
$P_I(k)$	Probability distribution of ingredients appearing in k recipes
$P_c(k)$	Complementary cumulative degree distribution
P_i^C	Prevalence of ingredient ***i*** in cuisine C