**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**NOVEL ALGORITHMS FOR INFERRING
GRANGER CAUSALITY USING BIOLOGICAL DATA**

MOHAMMAD SHAHERYAR FURQAN

**SCHOOL OF ELECTRICAL AND ELECTRONIC
ENGINEERING**

2017

STATEMENT OF ORIGINALITY

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

Date

Mohammad Shaheryar Furqan

**NOVEL ALGORITHMS FOR INFERRING
GRANGER CAUSALITY USING BIOLOGICAL DATA**

MOHAMMAD SHAHERYAR FURQAN

School of Electrical and Electronic Engineering

A thesis submitted to Nanyang Technological University in
partial fulfilment of the requirements for the degree of
Doctor of Philosophy

2017

ACKNOWLEDGEMENTS

Undertaking this Ph.D. has been a truly life-changing experience for me and requires saluting my dear and near ones by admiring them as a source of inspiration and motivation.

It is my utmost duty to show gratitude towards my advisor, Associate Professor Mohammed Yakoob Siyal, as it is under his supervision that this work came into existence. His sage advice, insightful criticisms, and patient encouragement from the initial to the final level aided my Ph.D. study in countless ways. His advice on both research as well as on my career has been invaluable. He had confidence in me when I doubted myself. It is his multidimensional skillful knowledge and leadership that assisted me in writing this thesis and journal and conference papers.

I am also indebted to Dr. Ghazanfar Monir, who convinced me during our many discussions that I should pursue my doctoral degree. I thank him for astounding editing of some research papers, valuable advice, constructive criticism and insightful feedback.

I would like to thank my parents Muhammad Furqan and Talat Furqan for being supportive of my decision to undertake this pursuit and for always believing in me and encouraging me to follow my dreams. Their prayers, encouragement, unflagging help, and love, in every unequivocal way, had made this endeavor a success story. Moreover, I want to thank my sister Dr. Husna Unnum Furqan for

listening to my complaints and frustrations, and for believing in me. She helped me enormously, especially with the final formatting and editing of my thesis.

I would also like to extend my appreciation to the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore for awarding me the Research Scholarship and providing me with excellent research facilities which enabled me to focus on research without other distractions. Likewise, I want to thank Laboratory staff of Information System Research Laboratory for providing timely support whenever it was required.

Towards the end of this work, I candidly thank everyone who had contributed, through their support and their recommendations, towards realization and successfully completing my Ph.D. study.

TABLE OF CONTENTS

Acknowledgements	iii
Table of Contents	v
Summary	ix
List of figures	xi
List of tables	xvii
List of abbreviations	xix
1 Introduction	1
1.1 Overview	1
<i>1.1.1 The brain networks</i>	<i>1</i>
1.1.1.1 Structural network	2
1.1.1.2 Functional network	3
1.1.1.3 Effective network	3
1.1.1.4 Importance of inference of effective connectivity	3
<i>1.1.2 Gene networks</i>	<i>4</i>
1.2 Motivation of the Study	6
1.3 Granger Causality	7
1.4 Contribution of the Thesis	10
1.5 Outline of the Thesis	11

2	Granger causality	13
2.1	Background	13
2.1.1	<i>Bivariate autoregressive models</i>	14
2.1.2	<i>Multivariate autoregressive models</i>	16
2.2	Model Selection.....	18
2.2.1	<i>Hold-out method</i>	19
2.2.2	<i>K-fold cross-validation</i>	20
2.2.3	<i>Leave-one-out cross-validation</i>	21
2.2.4	<i>Akaike information criterion</i>	21
2.2.5	<i>Bayes information criterion</i>	22
2.2.6	<i>Bootstrapping</i>	23
2.3	Estimation of Weight Matrix W and Their Limitations	24
3	Random Forest Granger causality	27
3.1	Background	27
3.1.1	<i>Tree-based algorithms</i>	27
3.1.2	<i>Bootstrap aggregation</i>	29
3.1.3	<i>Random forest</i>	29
3.2	Performance Evaluation of Random Forest Granger Causality	32
3.2.1	<i>Experimentation</i>	32

3.2.2	<i>Review of datasets</i>	33
3.2.2.1	Simulated dataset	33
3.2.2.2	Real dataset	36
3.3	Results and Discussion.....	39
3.4	Summary	51
4	Copula Granger causality	53
4.1	Introduction	53
4.2	Limitations of Existing Work.....	55
4.3	Proposed Elastic-net Copula Granger Causality	56
4.4	Proposed Random Forest Copula Granger Causality.....	57
4.5	Performance of Proposed Variants of Copula Granger Causality..	59
4.5.1	<i>Experimentation</i>	59
4.6	Results and Discussion.....	60
4.7	Summary	69
5	Bi-directional Granger causality.....	71
5.1	Introduction	71
5.2	Naïve Forward-Backward Granger Causality	71
5.3	Forward Backward Pairwise Granger Causality	73
5.3.1	<i>Pairwise Granger causality</i>	73
5.4	Bi-directional Random Forest Granger Causality	75

5.5	Performance Evaluation	76
5.5.1	<i>Simulated dataset</i>	77
5.5.2	<i>Real biological dataset</i>	77
5.6	Results and Discussions	77
5.6.1	<i>Simulated dataset</i>	77
5.6.2	<i>Real dataset</i>	80
5.7	Summary	83
6	Conclusion and Future Work	85
6.1	Conclusion	85
6.2	Future Work	87
	List of Publications	89
	Journal paper	89
	Conference papers	90
	References	91

SUMMARY

Biological network diagrams provide a natural means to characterize the association between biological entities such as genes, proteins, or brain regions. The understanding of these biological networks provides a range of information from systematic behaviours to disease susceptibility and its treatment. Currently, researchers from diverse research backgrounds are trying to model biological networks using various analytical modalities to improve understanding and prediction of biological networks. However, with the advancement of technology, the inference of biological networks from high-throughput data has received immense consideration throughout the last decade, and is a major area of research in systems biology.

One of the conventional methods used for inferring biological networks from time series data is Granger causality, which analyses the causal link between the variables and can be used to map the associations in the form of a network diagram using graph theory. However, when using high-throughput data, current implementations of Granger causality face limitations, especially while handling high dimensionality and non-linear relationships among correlated variables.

In order to minimize these impediments, we have proposed three heuristic techniques, and two variants of bi-directional Granger causality. We call the heuristic techniques random Forest Granger causality (RFGC), Elastic-net copula Granger causality (ECGC) and random forest copula Granger causality. The bi-directional

Granger causality methods include forward backward pairwise Granger causality (FCPGC) and bi-directional random forest Granger causality (BRFGC). All these techniques are capable of handling both high- and low-dimensional data. The effectiveness of these methods is demonstrated through extensive experimentation using simulated and real datasets. We also used these methods on two real world biological applications: in mapping a gene network, commonly involved in cancer using a HeLa cancer cell dataset; and in an effective brain network, normally involved in human deductive reasoning.

LIST OF FIGURES

FIGURE 1-1: DISTINCTIONS OF BRAIN CONNECTIVITY.	2
FIGURE 1-2: NODE AND EDGE REPRESENTATION	6
FIGURE 2-1: VISUALIZATION OF THE HOLD-OUT METHOD	19
FIGURE 2-2: VISUALIZATION OF FIVE-FOLD CROSS-VALIDATION	20
FIGURE 2-3: VISUALIZATION OF THE LEAVE-ONE-OUT METHOD.....	21
FIGURE 2-4: VISUALIZATION OF BOOTSTRAPPING	23
FIGURE 2-5: FLOW CHART OF GRANGER CAUSALITY EVALUATION USING LASSO AND ELASTIC-NET REGULARIZATION	26
FIGURE 3-1: BASIC TREE STRUCTURE	28
FIGURE 3-2: FLOW CHART OF GRANGER CAUSALITY EVALUATION USING RANDOM FOREST REGULARIZATION	32
FIGURE 3-3: INFLUENCE GRAPH FOR SIMULATED DATASET 1.	35
FIGURE 3-4: INFLUENCE GRAPH FOR SIMULATED DATASET 2.	35
FIGURE 3-5: INFORMATION LOSS CRITERION PLOT USING AIC AND BIC FOR MODEL SELECTION FOR SIMULATED DATASET 1.	42

FIGURE 3-6: INFORMATION LOSS CRITERION PLOT USING AIC AND BIC FOR MODEL SELECTION FOR SIMULATED DATASET 2.	42
FIGURE 3-7: INFORMATION LOSS CRITERION PLOT USING AIC AND BIC FOR MODEL SELECTION FOR HeLa CELL DATASET	43
FIGURE 3-8: INFORMATION LOSS CRITERION PLOT USING AIC AND BIC FOR MODEL SELECTION FOR STARPLUS DATASET.....	43
FIGURE 3-9: RESULTS OF THE COMPARATIVE STUDY OF LASSO GANGER CAUSALITY, ELASTIC-NET GRANGER CAUSALITY, RANDOM FOREST GRANGER CAUSALITY, AND STANDARD OLS GRANGER CAUSALITY ON SIMULATED DATASET 1.	46
FIGURE 3-10: RESULTS OF COMPARATIVE STUDY OF LASSO GANGER CAUSALITY, ELASTIC-NET GRANGER CAUSALITY, RANDOM FOREST GRANGER CAUSALITY, AND STANDARD OLS GRANGER CAUSALITY ON SIMULATED DATASET 2.	46
FIGURE 3-11: DIRECTED GRAPH FOR REAL STARPLUS DATASET USING SEVEN ROIs USING RANDOM FOREST GRANGER CAUSALITY.	47
FIGURE 3-12: DIRECTED MAP FOR NINETEEN PRE-SELECTED GENES USING RANDOM FOREST GRANGER CAUSALITY	47
FIGURE 4-1: INFORMATION LOSS VS MODEL ORDER PLOT USING AIC AND BIC FOR MODEL SELECTION FOR SIMULATED DATASET 2.	61
FIGURE 4-2: INFORMATION LOSS VS MODEL ORDER PLOT USING AIC AND BIC FOR MODEL SELECTION FOR SIMULATED DATASET 1.	61

FIGURE 4-3: INFORMATION LOSS VS MODEL ORDER PLOT USING AIC AND BIC FOR MODEL SELECTION FOR HELA CANCER CELL DATASET.	62
FIGURE 4-4: INFORMATION LOSS VS MODEL ORDER PLOT USING AIC AND BIC FOR MODEL SELECTION FOR FMRI STARPLUS DATASET.	62
FIGURE 4-5: RESULTS FOR COMPARATIVE STUDY OF ELASTIC-NET COPULA GRANGER CAUSALITY, LASSO COPULA GRANGER CAUSALITY, RANDOM FOREST COPULA GRANGER CAUSALITY, AND STANDARD OLS COPULA GRANGER CAUSALITY ON SIMULATED DATA SET 1.	65
FIGURE 4-6: RESULTS FOR COMPARATIVE STUDY OF ELASTIC-NET COPULA GRANGER CAUSALITY, LASSO COPULA GRANGER CAUSALITY, RANDOM FOREST COPULA GRANGER CAUSALITY, AND STANDARD OLS COPULA GRANGER CAUSALITY ON SIMULATED DATA SET 2.	65
FIGURE 4-7: BRAIN NETWORK INVOLVED IN DEDUCTIVE REASONING USING ELASTIC- NET COPULA GRANGER CAUSALITY.	66
FIGURE 4-8: BRAIN NETWORK INVOLVED IN DEDUCTIVE REASONING USING RANDOM FOREST COPULA GRANGER CAUSALITY.	66
FIGURE 5-1: DEPICTION OF THE CONCEPT BEHIND BI-DIRECTIONAL GRANGER CAUSALITY.....	72
FIGURE 5-2: PRECISION AND FALSE DISCOVERY RATE RESULTS FOR A COMPARATIVE STUDY AMONG LASSO, ELASTIC-NET, PAIRWISE, AND STANDARD OLS – BASED GRANGER CAUSALITY.	74

FIGURE 5-3: ACCURACY AND F1-SCORE RESULTS FOR A COMPARATIVE STUDY AMONG LASSO, ELASTIC-NET, PAIRWISE, AND STANDARD OLS – BASED GRANGER CAUSALITY.....	74
FIGURE 5-4: RESULTS OF A COMPARATIVE STUDY BETWEEN FORWARD BACKWARD PAIRWISE GRANGER CAUSALITY AND LASSO-BASED FORWARD BACKWARD GRANGER CAUSALITY.....	78
FIGURE 5-5: RESULTS OF A COMPARATIVE STUDY OF BI-DIRECTIONAL RANDOM FOREST GRANGER CAUSALITY VS. LASSO-BASED FORWARD BACKWARD GRANGER CAUSALITY.....	79
FIGURE 5-6: RESULTS OF A COMPARATIVE STUDY OF STANDARD OLS-BASED FORWARD BACKWARD GRANGER CAUSALITY VS. LASSO-BASED FORWARD BACKWARD GRANGER CAUSALITY.....	79
FIGURE 5-7: COMBINED RESULTS OF SIMULATED STUDY ON BI-DIRECTIONAL GRANGER CAUSALITY.....	80
FIGURE 5-8: INFERRED GENE NETWORK FOR 19 PRESELECTED GENES USING FORWARD BACKWARD PAIRWISE GRANGER CAUSALITY.	80
FIGURE 5-9: GENE NETWORK MAP FOUND USING BI-DIRECTIONAL RANDOM FOREST GRANGER CAUSALITY.....	81
FIGURE 5-10: EFFECTIVE BRAIN CONNECTIVITY MAP USING FORWARD BACKWARD PAIRWISE GRANGER CAUSALITY.	81

FIGURE 5-11: EFFECTIVE BRAIN CONNECTIVITY MAP USING BI-DIRECTIONAL RANDOM
FOREST GRANGER CAUSALITY. 82

LIST OF TABLES

TABLE 1: RESULTS OF FOURTEEN-VARIABLE SCENARIO FOR COMPARATIVE STUDY OF LASSO GRANGER CAUSALITY, ELASTIC-NET GRANGER CAUSALITY, RANDOM FOREST GRANGER CAUSALITY, AND STANDARD OLS GRANGER CAUSALITY....	44
TABLE 2: RESULTS OF FIVE-VARIABLE SCENARIO FOR COMPARATIVE STUDY OF LASSO GRANGER CAUSALITY, ELASTIC-NET GRANGER CAUSALITY, RANDOM FOREST GRANGER CAUSALITY, AND STANDARD OLS GRANGER CAUSALITY.....	45
TABLE 3: TOP 20 SIGNIFICANT GENE INTERACTIONS USING LASSO REGULARIZED GRANGER CAUSALITY, ELASTIC-NET REGULARIZED GRANGER CAUSALITY AND RANDOM FOREST GRANGER CAUSALITY	48
TABLE 4: RESULTS FOR COMPARATIVE STUDY OF ELASTIC-NET COPULA GRANGER CAUSALITY, LASSO COPULA GRANGER CAUSALITY, RANDOM FOREST COPULA GRANGER CAUSALITY, AND STANDARD OLS COPULA GRANGER CAUSALITY ON SIMULATED DATA SET 1.	63
TABLE 5: RESULTS FOR COMPARATIVE STUDY OF ELASTIC-NET COPULA GRANGER CAUSALITY, LASSO COPULA GRANGER CAUSALITY, AND STANDARD OLS COPULA GRANGER CAUSALITY ON SIMULATED DATA SET 2.....	64

TABLE 6 : TOP 20 SIGNIFICANT GENE INTERACTIONS USING LASSO COPULA GRANGER CAUSALITY, ELASTIC-NET COPULA GRANGER CAUSALITY, RANDOM FOREST COPULA GRANGER CAUSALITY, AND STANDARD OLS BASED COPULA GRANGER CAUSALITY.....	67
--	----

LIST OF ABBREVIATIONS

BioGRID	Biological general Repository for Interaction Datasets
BRFGC	Bi-directional Random Forest Granger causality
CALC	Calcarine sulcus
cRNA	cyclic Ribonucleic Acid
DNA	Deoxyribonucleic Acid
ECGC	Elastic-net copula Granger causality
EEG	Electroencephalograph
E-NET	Elastic-net
FBGC	Forward Backward Granger causality
FBPGC	Forward Backward Pairwise Granger causality
FDR	False Discovery Rate
fMRI	Functional Magnetic Resonance Imaging
GC	Granger causality
GCCA	Granger Causal Connectivity Analysis
GNPN	Gaussian
LASSO	Least Absolute Shrinkage and Selection Operator
LDLPFC	Left dorsolateral prefrontal cortex

LFEF	Left frontal eye fields
LIPL	Left inferior parietal lobule
LIFG	Left inferior frontal gyrus
LIPS	Left intraparietal sulcus
LIT	Left inferior temporal lobule
LOPER	Left opercularis
LPPREC	Left posterior precentral sulcus
LSPL	Left superior parietal lobule
LSGA	Left supramarginal gyrus
LTRIA	Left triangularis
LT	Left temporal Lobule
mRNA	messenger Ribonucleic Acid
OLS	ordinary least square
RDLPFC	Right dorsolateral prefrontal cortex
RF	Random forest
RFEF	Right frontal eye fields
RFGC	Random Forest Granger causality
RIPL	Right inferior parietal lobule
RIPS	Right intraparietal sulcus
RIT	Right inferior temporal lobule

ROPER	Right opercularis
RPPREC	Right posterior precentral sulcus
RSPL	Right superior parietal lobule
RSGA	Right supramarginal gyrus
RTRIA	Right trianularis
RT	Right temporal Lobule
ROI	Region of Interest
RSS	Residual Sum of Square
SLEP	Sparse Learning with Efficient Projections
SMA	Supplementary motor areas
VAR	Vector Auto-Regressive

1 INTRODUCTION

1.1 Overview

Components of biological systems such as genes, proteins, metabolites, and neurons interact with each other to carry out vital functions in a specific sequence. These interactions constitute a biological network. The analysis of biological networks provides a mathematical foundation for researchers to gain deeper insights into functions, mechanisms, and principles of biological systems [1].

Of the various biological networks, this thesis has focused on two significant and vital networks: the brain and gene networks.

1.1.1 The brain networks

The brain is a complex and unpredictable arrangement of interconnected cells which provides the brain with an essential structure that is commonly known to us, and is the basis for intra and intercommunication between distinctive brain areas [2-4]. These connections constitute a brain network. Studies related to various brain networks are inspired by the notion that the brain functions cannot be exclusively ascribed to discrete regions and connections [4-6], but rather can be attributed to activities of the network topology [7]. Thus, researchers are motivated to develop means to extract relevant information related to networks to explain the mechanisms of information processing by neurons and neural networks.

The term brain network can denote several different and interrelated aspects of brain organization [8]. Of these, fundamental distinctions exist between structural, functional, and effective networks [9] as shown in Figure 1-1 (taken from [10]).

Figure 1-1: Distinctions of brain connectivity.

1.1.1.1 Structural network

The structural network, also known as the anatomical network, refers to the anatomical/physical connections in the human brain [9]. It pertains to the physical links of fibre tracks that are associated with the pool of neurons across different functional brain regions.

1.1.1.2 Functional network

A functional network is the network of spatially distributed structural regions that are functionally connected to perform a specific function. These functionally connected regions can be mapped by using statistical concepts that capture the deviation of statistical independence [9]. This statistical independence can be estimated using statistical measures of correlation, covariance, spectral coherence etc.

1.1.1.3 Effective network

An effective network can be perceived as a combination of structural and functional networks [11], and depends directly on tasks and contexts. Thus, it provides a mechanistic account of the cause of inter-regional interactions that may explain the emergence of a functional pattern.

1.1.1.4 Importance of inference of effective connectivity

During the recording of electrophysiological brain activities to find the effective network, stimulus-locked transients produce confounding effects. These effects are not due to direct stimulation of the structural network. In neuroimaging, stimulus-locked transients do not significantly affect the recording, as different primary sensory areas are bi-directionally connected functionally and structurally with each other. Each functional network is converted into an effective network due to multi-directional signals from the relay center (the thalamus) and other related functional areas (multiple anatomical connections) [12].

Thus, in this thesis, we have focused our study on the visualization of effective brain connectivity networks.

1.1.2 Gene networks

The human cell is one of the most complex constructs that comprises different organic and inorganic substances. Organic cell components include different proteins coded by the DNA molecules located in the nucleus. DNA is considered the key to life as it contains all information related to inheritance transferred from generation to generation. These pieces of information are called genes, and are made up of a sequence of triplet nucleotides on the DNA. These sequences of triplets depict the sequence of amino acids of specific proteins in the form of a genetic code.

Genes can be divided into two broad categories, that is, constitutive and regulatory [13]. Constitutive genes code for all the functional and structural proteins required at variable periods, under the influence of environmental changes during a lifespan. In contrast, regulatory genes are responsible for controlling the time and duration of a constitutive gene expression. For example, enzymes of metabolic activities are increased or decreased according to the requirements of the body. Under abnormal conditions, these regulatory genes also help maneuver the expression or suppression of constitutive genes, which are not normally expressed.

A regulatory gene can affect a constitutive gene located in its vicinity or at a distance, or affect a genetic code present on other chromosomes. Therefore, an understanding of the communication mechanism of these two groups of genes will enable us to understand the functioning of the human cell. The interaction between constitutive and regulatory genes is called the gene network.

The goal of gene network analysis is to examine and predict the gene-gene interaction individually or as functional blocks. In order to examine any level of

interaction, we need to know two things about the network: its structure and the flow of information between unit blocks. An understanding of the connections and interactions of different genes can prove useful in several ways, from understanding disease susceptibility to its treatment, and from understanding gene behavior to finding ways for effective and efficient delivery of drugs.

A gene regulatory network is initiated by the production of a specific mRNA for a specific protein by a regulatory gene under the influence of biochemical and environmental factors. The resultant protein formed by the decoding of regulatory genes, called a transcriptional factor, will cause activation of a constitutive gene, which transcribes a specific mRNA for a specific protein for specific biological functions. This entire cascade of expression is called the gene regulatory network [1]. The study of this cascade is a dynamic research area in the fields of biology, genometrics, bioinformatics, and other related fields.

A typical genetic network is composed of nodes and edges. The nodes of a genetic network are represented by genes that produce proteins and its metabolites, whereas the proteins and metabolites that interact with other nodes constitute the edges that are produced by the transcription of DNA into RNA by specific nodes. This concept of genetic network is explained in Figure 1-2 (taken from [14]).

Figure 1-2: Node and Edge Representation

The genetic network shown in Figure 1-2 depicts a network in which *Gene A* transcribes mRNA of a protein that regulates the expression of *Gene B*, *Gene C*, and itself. The protein and its metabolites that regulate the expression of other genes can be observed at the transcriptional level and the entire activated genetic network can be simply summarized as the dot-arrow diagram as shown at the influential level.

In this thesis, we have focused our study on the genetic networks that are presumably involved in the cancer cell division cycle.

1.2 Motivation of the Study

In the current age of medicine, researchers from different areas are working together to decipher the functionality of various biological networks in an effort to diagnose and treat diseases at the core level. Thus, understanding the structure and intra- and inter-communications that exist between the affected constituents (genes or regions of brain) would render the healing process more proficient and faster.

However, with advancement in technology, techniques like DNA microarray and functional magnetic resonance imaging (fMRI) yield high throughput biological data by simultaneously analysing multiple variables. The availability of such high-dimensional data motivates researchers to explore new, improved techniques that can utilize high dimensional temporal data for reliable and accurate understanding and prediction of biological systems [15].

Recently, biological networks have been analysed using various mathematical concepts, including information theory [16], control theory [17], Bayesian statistics [18], and ordinary differential equations [19, 20]. However, a correlation-based algorithm like Granger causality attracts more researchers [21, 22].

1.3 Granger Causality

The concept of causal influence can be traced back to 1956 when Wiener [23] conceived the idea that if the inclusion of information from one time series can improve the prediction of another time series, the second series has a causal influence on the first. After more than a decade, in 1969, the concept was practically formalized by Granger [21], who used the concept to analyse the causal interaction between financial time-series data. Since then, the concept of Granger causality has been discussed and refined by many researchers, including Pierce and Haugh [24] and Geweke [25], while Lütkepohl [26] and Thostheim [27] have vectorized its formalization. Later, the concept of Granger causality was introduced in bioinformatics to explore structural and functional aspects of different biological systems, such as gene networks [28-34], and brain networks [35-39].

Since the concept of Granger causality does not rely on any specific modelling techniques, most applications in bioinformatics use vector autoregressive (VAR) models to implement and interpret time-series data [21, 40, 41]. The use of VAR models has proved advantageous, especially for describing processes that are composed of locally interacting components. Moreover, the statistical concepts behind the estimation of VAR models using observational data and its computational algorithms are well understood. Some VAR models have been employed successfully in several areas, such as economics [21, 42, 43], neuroscience [44-48], and, more recently, system biology [29, 49-53].

Other than VAR model variants, other non-model variants, like those using partial correlations, can be utilized for the general identification of Granger causality [54]. These partial correlation methods are especially effective when dealing with data that have unobserved (latent) variables [55].

In the past, several bioinformatics studies have addressed different aspects of Granger causality [56]. For instance, it is challenging or sometimes unfeasible to incorporate all variables in the model. Thus, Guo *et al.* [57] developed a method that reduces the influence of latent variables in both time and frequency domains. Fujita *et al.* [58] developed a VAR model that incorporates microarray measurement errors, yielding more precise p-values and a more precise model.

Another important aspect to address is the use of Granger causality for analysing non-linear relationships among multiple time series. As discussed earlier in the introduction, the standard formulation of Granger causality by Granger [21] was developed to analyse linear relationships among several time series. However,

its use to observe non-linear relationships might not identify the correct interaction patterns.

In the past, researchers like Priestley [59], Freiwald *et al.* [47], and Schreiber [60] discussed ways to analyse the interdependence between variables in non-linear systems. However, the works of Arnhold *et al.* [61], Quain Quiroga *et al.* [62, 63], Bhattacharya *et al.* [64], and Breakspear *et al.* [65] use the concept of delay-coordinate embedding reconstruction of phase space. According to this concept, the neighbourhood points in the reconstructed space are used to predict the dynamic behaviour between time series. These methods discuss different criteria to quantify the accuracy of prediction. However, these methods work mainly to detect strong interactions. In 2001, an extension to these methods, a mixed-state prediction method to detect weak interactions, was proposed by Wiesenfeldt *et al.* [66]. However, all these methods use a zero-order predictor, which takes the mean or weighted mean as the prediction value, and do not properly account for the flow of time.

Despite improvements in recent years, Granger causality still faces limitations regarding the handling of high dimensionality and non-linear relationships among variables. Addressing these two problems would improve our understanding of information flow in biological systems, and would provide additional insights into cellular processes. This would ultimately lead researchers to better comprehension of disease mechanisms and identification of potential drug targets for better treatment of diseases at their source.

Therefore, in this work, we propose some heuristic implementations of Granger causality to resolve the impediments of high dimensionality and non-linear

relationships among variables. These new techniques can be used in various biological applications, including deciphering the functionality of genetic code present in DNA molecules and for mapping of effective brain connectivity. This research provides a visualization tool for the medical research community to help in the diagnosis and treatment of diseases at their core by understanding the structure and intra- and inter-communications networks that exist between its constituents (genes or brain regions). Thus, this study helps researchers to make the healing process faster and more proficient.

1.4 Contribution of the Thesis

The contributions of this thesis are as follows.

Contribution 1: We propose a technique called random forest Granger causality. This technique is capable of handling both high- and low- dimensional data effectively using random forest as a regularization technique for evaluating Granger causality. The effectiveness of the proposed method is demonstrated using extensive experimentation on simulated data. This method is also used for the inference of gene networks, and mapping of effective brain connectivity on real biological datasets.

Contribution 2: We propose two variants of copula-based Granger causality called elastic-net copula Granger causality and random forest copula Granger causality. These techniques are also capable of effectively handling both high- and low-dimensional data, and can extract both linear and non-linear relationships. Elastic-net copula Granger causality uses elastic-net regularization to handle data dimensionality, and uses copulas to handle linearity issues, whereas random forest copula Granger causality uses random forest as a regularization technique to evaluate

Granger causality. The effectiveness of these methods is discussed in Chapter 4, using experimentations on simulated datasets. It is also used on real biological datasets to map gene and brain networks.

Contribution 3: The third contribution is based on the concept of isotropy of time, which suggests that the direction of time is irrelevant. Using this concept, we propose two heuristic variants of bi-directional Granger causality analysis techniques, namely, bi-directional random forest Granger causality and forward backward pairwise Granger causality. These techniques evaluate Granger causality as a linear weight sum of results obtained from standard time-stamped Granger causality analysis and the results obtained by reusing the time-series data in reverse order. We compared the effectiveness of the proposed methods against the existing least absolute shrinkage and selection operator (lasso)-based method using simulated datasets and then use them on real datasets to map gene and brain networks involved in cancer and deductive reasoning, respectively.

1.5 Outline of the Thesis

In addition to this introductory Chapter 1, this thesis consists of five more chapters. Chapter 2 discusses the existing implementations of Granger causality. In Chapter 3, we discuss the first proposed method, called random forest Granger causality, evaluate its performance against existing methods, and apply it to real biological datasets. Chapter 4 discusses the second contribution of the thesis by introducing elastic-net Granger causality and random forest copula Granger causality, their performance evaluations against existing methods, and their application to real biological datasets. Chapter 5 covers the third contribution of the thesis by

introducing two variants of bi-directional Granger causality techniques namely, bi-directional random forest Granger causality and forward backward pairwise Granger causality. This chapter also includes a performance evaluation of the proposed methods using simulated datasets and their application to real biological datasets. Chapter 6 provides a summary of the thesis with insights on possible future work.

2 GRANGER CAUSALITY

2.1 Background

In the last few decades, the concept of Granger causality has been used by many researchers to map causal interactions among time-series data. The concept of Granger causality first surfaced in 1969 by Granger [21], who used the notion to contemplate the causative effect of different financial time series. Later, the same concept was used in bioinformatics to analyse various aspects, such as functional and effective connectivity of the brain, using fMRI data [35, 39] and in genetics for analysis of gene networks [29, 30].

In Granger causality, variable X Granger causes Y if using the past values of both variables X and Y to predict the future of X leads to a better prediction rather than using only the previous X values. Thus, the concept of Granger causality can provide a mechanistic way to account for the causes of the inter-variable interactions, and helps to analyse the direction of influences between variables.

The concept of Granger causality can be mathematically implemented using different modeling techniques, like the autoregressive Baba-Engle-Kraft-Kroner (BEKK) [67], VAR, and vector autoregressive moving average (VARMA) models, as discussed in detail in Lutkepohl [26]. However, in bioinformatics, the VAR method is widely used because of its ease of implementation and interpretation [9, 36, 37].

The basic VAR-based implementation of Granger causality can be formulized using a bivariate VAR model, and then the same concept can be extended to the multivariate scenario.

2.1.1 Bivariate autoregressive models

As discussed by Tam *et al.* [30], we consider a two-variable system that can be articulated using time series $x_1(t)$ and $x_2(t)$ of length T , where $t = 1, 2, \dots, T$. The autoregressive model can be mathematically described as

$$x_1(t) = \sum_{i=1}^p (a_{11,i}x_1(t-i) + a_{12,i}x_2(t-i)) + \varepsilon_1(t)$$

$$x_2(t) = \sum_{i=1}^p (a_{21,i}x_1(t-i) + a_{22,i}x_2(t-i)) + \varepsilon_2(t)$$

where p is the model order and $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are independent and identically distributed white noises with $E(\varepsilon_1(t)) = E(\varepsilon_2(t)) = 0$, $E(\varepsilon_1(t)\varepsilon_1(t)') = E(\varepsilon_2(t)\varepsilon_2(t)') = \varepsilon$ and $E(\varepsilon_1(t)\varepsilon_2(t)') = E(\varepsilon_2(t)\varepsilon_1(t)') = 0$.

For $t = p + 1, \dots, T$, there are $m = T - p$ pairs of equations. This set of equations for a bivariate system can be reformulated in matrix form as follows:

$$Y = XW + \Sigma$$

where

$$\begin{matrix} Y \\ (m \times 2) \end{matrix} = \begin{bmatrix} x_1(p+1) & x_2(p+1) \\ x_1(p+2) & x_2(p+2) \\ \vdots & \vdots \\ x_1(T) & x_2(T) \end{bmatrix}$$

$$X_{(m \times 2p)} = \begin{bmatrix} x_1(p) & x_2(p) & x_1(p-1) & x_2(p-1) & \cdots & x_1(1) & x_2(1) \\ x_1(p+1) & x_2(p+1) & x_1(p) & x_2(p) & \cdots & x_1(2) & x_2(2) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1(T-1) & x_2(T-1) & x_1(T-2) & x_2(T-2) & \cdots & x_1(T-p) & x_2(T-p) \end{bmatrix}$$

$$W_{(2p \times 2)} = \begin{bmatrix} A'_1 \\ A'_2 \\ \vdots \\ A'_p \end{bmatrix} \text{ where } A'_i = \begin{bmatrix} a_{11,i} & a_{21,i} \\ a_{12,i} & a_{22,i} \end{bmatrix}, i = 1, 2, \dots, p;$$

$$\Sigma_{(m \times 2)} = \begin{bmatrix} \varepsilon_1(p+1) & \varepsilon_2(p+1) \\ \varepsilon_1(p+2) & \varepsilon_2(p+2) \\ \vdots & \vdots \\ \varepsilon_1(T) & \varepsilon_2(T) \end{bmatrix}$$

The unknown weight matrix W can be estimated using different techniques, such as Equation 1, which uses ordinary least squares (OLS):

$$\hat{W} = (X'X)^{-1}X'Y \quad (1)$$

Once all matrixes are calculated, the prediction error (residual sum of squares) is calculated as:

$$RSS_1 = (y - X\hat{W})(y - X\hat{W})' \quad (2)$$

where y and $\hat{W}$ are the initial columns of Y and $\hat{W}$, respectively and RSS_1 represents $RSS_{un-restricted}$, as both variables have influenced the value.

The basic concept of Granger causality is based on the premise that prediction using two or more variables should yield better results than using an individual variable. Therefore, to see the impact of an individual variable, $RSS_{restricted}$ is calculated using time series having a single variable. For example, to observe whether x_2 Granger causes x_1 , the following AR models are used:

$$x_1(t) = \sum_{i=1}^p a_{11,i} x_1(t-i) + \varepsilon_1(t)$$

$$x_2(t) = \sum_{i=1}^p a_{22,i} x_2(t-i) + \varepsilon_2(t)$$

Moreover, the residual sum of squares RSS_{21} is calculated using Equation (2), which yields

$$RSS_{21} = (y - X\hat{w})(y - X\hat{w})'$$

Using the values of the residual sum of squares, the F-statistic is calculated as follows:

$$F = \frac{(RSS_{21} - RSS_1)/p}{RSS_1/(m-2p)}.$$

For this F-test, the null hypothesis is the non-existence of causality and the F-statistic follows an F distribution with p and $m - 2p$ degrees of freedom for the numerator and denominator, respectively.

2.1.2 Multivariate autoregressive models

In order to extend the bivariate implementation to the multivariate case, that is, when the *number of variables* > 2 , conditional Granger causality is evaluated instead of simple Granger causality [26, 30, 68, 69]. In conditional Granger causality, each variable is left out, in sequence, and its effect on the residual sum of squares is evaluated. Evidence of Granger causality between variables is detected if the accumulation of variables attenuates the residual sum of squares for predictions.

In order to understand multivariate implementation as described by Yao *et al.* [52] and Siyal *et al.* [69], consider a scenario of n -variables, where each variable has an equal number of total observational time points T . This system can be written as a p -order VAR model in the following form:

$$x(t) = \sum_{i=1}^p A_i x(t-i) + \varepsilon(t); \quad \text{where } t = p+1, \dots, T$$

where $x(t)$ is an $n \times 1$ vector comprising values of all the variables at time point t , A_i is the matrix of parameters of the AR model, and $\varepsilon(t)$ is the $n \times 1$ vector of residuals. Since there are $m = T - p$ pairs of equations, its matrix representation is as follows.

$$Y = XW + \Sigma$$

where

$$Y_{(m \times n)} = \begin{bmatrix} x'_{p+1} \\ x'_{p+2} \\ \vdots \\ x'_T \end{bmatrix}; \quad X_{(m \times np)} = \begin{bmatrix} x'_p & x'_{p-1} & \dots & x'_1 \\ x'_{p+1} & x'_p & \dots & x'_2 \\ \vdots & \vdots & \ddots & \vdots \\ x'_{T-1} & x'_{T-2} & \dots & x'_{T-p} \end{bmatrix}$$

$$W_{(np \times n)} = \begin{bmatrix} A'_1 \\ A'_2 \\ \vdots \\ A'_p \end{bmatrix}; \quad \Sigma_{(m \times n)} = \begin{bmatrix} \varepsilon'_{p+1} \\ \varepsilon'_{p+2} \\ \vdots \\ \varepsilon'_T \end{bmatrix}$$

In estimating the matrix W , the standard method shown in Equation (1) is used, provided the matrix $X'X$ is invertible and $T \geq (n+1)p$. The unrestricted RSS_i is calculated using Equation (2) and the restricted RSS using *variable j* while

variable i is calculated by removing *variable j* and solving the linear regression with $n - 1$ variables using the above-mentioned procedure. It is denoted as RSS_{ji} .

Using restricted and unrestricted RSS values, the F-statistic is evaluated using Equation (3) under the null hypotheses that *variable j* does not Granger cause *variable i*:

$$F = \frac{(RSS_{ji} - RSS_i)/p}{RSS_i/(m - np)} \quad (3)$$

These F-values follows p and $m - 2p$ degrees of freedom for the numerator and denominator, respectively, and their corresponding p-values determine the presence of Granger causality.

2.2 Model Selection

One of the key steps for implementing Granger causality using the VAR model is to select the optimal model order. As it is difficult to find a single model that can incorporate all data characteristics and at the same time limit data complexity, we need to compare various models and select the optimal model that is simple, incorporates most of the data characteristics, and is generalizable to new data.

There are a number of methods in the literature that can compare the effectiveness of models, such as the hold-out method [70, 71], K-fold cross-validation and its variants [71], Akaike information criterion [72, 73], Bayesian information criterion (BIC) [72, 74], and bootstrapping [72, 75].

2.2.1 Hold-out method

The hold-out method is one of the basic model-validation techniques used in statistical analysis. It is implemented by randomly dividing the data into two sets (training and testing) while assuming the data are independently and identically distributed. The training set containing the major portion of data (i.e., 70-90 % of the dataset) is used to train the model, while the remaining 30-10 % of the data is left out to validate effectiveness of the predicted model. The reason for such uneven distribution of data in the training and testing datasets is to train a model that incorporates most characteristics of the data. The division of data is shown in Figure 2-1.

Figure 2-1: Visualization of the hold-out method

Computationally, the hold-out method is less expensive and easier to implement than other methods. However, its results are statistically less reliable because the use of a small validation dataset makes it difficult to estimate confidence intervals or variance information [71]. In practice, the hold-out method is used for datasets with a validation set large enough to ensure reliable statistical estimates or in conjunction with other validation methods.

2.2.2 K -fold cross-validation

K -fold cross-validation is an improved form of the hold-out method. Instead of dividing the data in two parts (testing and training), the data are divided into K – equal parts [71], as shown in Figure 2-2. Then, the hold-out method is implemented K times. In each repetition, $K - 1$ part of the data is used to train the model while the remaining part is used for validation. After K repetitions, the average error is computed across all K repetitions.

	Data				
	1	2	3	4	5
Fold 1	Train	Train	Train	Train	Test
Fold 2	Train	Train	Train	Test	Train
Fold 3	Train	Train	Test	Train	Train
Fold 4	Train	Test	Train	Train	Train
Fold 5	Test	Train	Train	Train	Train

Figure 2-2: Visualization of five-fold cross-validation

The advantage of K -fold cross-validation is that all data are used for testing and training. Thus, the resultant model incorporates almost all characteristics of the data. As K models are generated and mean values are used, the variance of resultant errors can be reduced by increasing the value of K . The disadvantage of K -fold cross-validation is that it is computationally expensive, as the method requires training and testing of K models.

2.2.3 Leave-one-out cross-validation

The leave-one-out method is a special case of K -fold cross-validation, in which K equals the total number of data points (n) [71]. This means n models are trained and tested, and each time only one data point is left out for testing. This variant of K -fold validation is an extreme cross-validation technique and its computational time increases as the size of the dataset increases. The visual representation of its implementation is shown in Figure 2-3.

Figure 2-3: Visualization of the leave-one-out method

2.2.4 Akaike information criterion

The Akaike information criterion (AIC) is one of the widely used model selection method. Formulated by Akaike in 1974 [73], it provides a way to estimate the model's loss of information (Kullback-Leibler information [76]) compared to the true model using maximum likelihood. This provides an objective way of determining the most parsimonious model among a set of models by penalizing the

addition of parameters. In other words, the AIC prefers a model that fits well and uses a minimum number of parameters. Its basic mathematical evaluation is as follows:

$$AIC = 2k - 2 \ln(P(x|\hat{\theta}, M))$$

where M is the statistical model, x is data, k is the number of variables (parameters) included in the model, and $P(x|\hat{\theta}, M)$ is the likelihood function for the model.

The advantages of using AIC include proficiency of model selection, ability to compare all possible models in a set, instead of a single model, and incorporation of most of the model uncertainty in the estimates and precision of parameters [77]. These advantages make AIC one of the widely used methods for model selection, not only for cases where there are large numbers of models but even for two-model cases [78]. However, the quality of model is highly dependent on the quality of data used to generate the model [79].

2.2.5 Bayes information criterion

In 1978, Schwarz proposed the BIC, a modified form of the AIC, as a model selection criterion [74]. Instead of using a scalar penalty function to reduce the effect of overfitting as in the AIC, Schwarz proposed using the logarithmic penalty function, which increases the bias toward complex models and filters out superfluous complex models that contain too many parameters. The BIC can be expressed mathematically as

$$BIC = \ln(n) k - 2 \ln(P(x|\hat{\theta}, M))$$

where M is the statistical model, x is data, k is the number of variables (parameters) included in the model, and $P(x|\hat{\theta}, M)$ is the likelihood function.

However, the BIC faces limitations when dealing with high-dimensional cases, that is, when the number of observations are greater than the number of variables (parameters) and when dealing with complex collections of models [80].

2.2.6 Bootstrapping

Bootstrapping is a resampling technique that is closely related to cross-validation [72]. In bootstrapping, multiple small datasets are generated by sampling the data from a large pool of given data [72, 75]. These sub-sampled datasets can then be used for model estimation, as all these subsets inherit the same data characteristics. Since multiple datasets are involved, the statistical significance of the estimates can be evaluated by calculating variance or confidence intervals of the estimated models.

The concept of bootstrap is inspired by another data resampling method called the jackknife [81], which is based on leave-one-out cross-validation. In the jackknife method, N -numbers of smaller data subsets are generated by sampling the data from given data without being replaced. In other words, if a data point is selected for one subset, the same data point cannot be selected for any other subset.

Figure 2-4: Visualization of bootstrapping

In contrast, bootstrapping divides the data through replacement while selecting the data for every subset at random as shown in Figure 2-4. The reason for replacing the data in bootstrapping is to preserve the original data distribution. If the data were resampled without replacement, the data distribution of the original dataset would keep changing with each data withdrawal and the resultant subsets would no longer mimic the data distribution of the original dataset [71, 72, 75].

The drawback of using the bootstrap method is that there is repetition of the same data in multiple datasets because of the replacement of data [71]. However, the effects of repetition can be reduced by

- 1) increasing the number of subsets used for estimation; and
- 2) splitting the data using the hold-out method and using different data for training and testing purposes [71].

2.3 Estimation of Weight Matrix W and Their Limitations

We know that the original application of Granger causality used OLS to calculate weight matrix W , which reduces the sum of the squared errors between real and projected values. There is satisfactory application of OLS methods for low-dimensional data, that is, when the number of variables is fewer than the number of observational time points. However, this method fails to handle high-dimensional data [82], because W is no longer full rank and $X^T X$ cannot be inverted [83]. Therefore, to handle high-dimensional data, OLS is replaced with other regularized methods [34, 84-87]. Regularization-based methods are more robust to high collinearity, tend to yield sparse solutions, and can account for variable groupings in

high-dimensional data-sets. Among the many regularization methods, lasso-based regularization has received wide acceptance [49, 88].

Rather than calculating the sum of the squared errors between real and expected values, lasso-based regularization minimizes the sum of the squared errors subject to a bound on the absolute value of the coefficient [89, 90] utilizing the following objective function:

$$\hat{A} = \underset{w}{\operatorname{argmin}} \|x(t) - XW\|^2 + \rho \|W\|_1$$

where ρ is the tuning parameter and $\|W\|_1 = \sum_{j=1}^p W_j$.

Subsequently, research discovered that lasso has difficulties when predictors are highly correlated with each other, even when the number of observed values is greater than the number of predictors. This limitation results from the dominance of ridge regression, as discussed comprehensively by Tibshirani [89], Hoerl and Kennard [91], and Zuo [92].

These confinements were later addressed by Zou and Hastie [82] through the addition of another constraint term, $\lambda \|W\|_2^2$. This refinement of the regularization function is called elastic-net (e-net) regularization [82], which utilizes the following objective function:

$$\hat{A} = \underset{w}{\operatorname{argmin}} \|x(t) - XW\|^2 + \lambda \|W\|_2^2 + \rho \|W\|_1$$

where λ and ρ are the tuning parameters, $\|W\|_2^2 = \sum_{j=1}^p W_j^2$, and $\|W\|_1 = \sum_{j=1}^p |W_j|$.

The lasso and elastic-net regularized implementation follow the steps as shown in Figure 2-5.

Figure 2-5: Flow chart of Granger causality evaluation using lasso and elastic-net regularization

The traditional regression techniques discussed in this chapter can be used for predicting weight matrixes, but due to high collinearity when using large datasets, these methods do not perform well in terms of precision, false detection rate, and other important measures. Therefore, in the following three chapters, we propose new and improved methods for evaluating Granger causality that are expected to outperform the existing methods in these measures.

3 RANDOM FOREST

GRANGER CAUSALITY

3.1 Background

In Chapter 2, we discussed the shortcomings of methods currently used for evaluating weight matrixes while analysing Granger casualty. In order to address these issues, in this chapter we propose using a tree-based method called random forest as a regularization technique in evaluating Granger causality.

3.1.1 Tree-based algorithms

The first tree-based algorithm was introduced in 1963 to analyse the multivariate non-additive effects of survey data by Morgan *et al.* [93]. Since then, tree-based methods have been enhanced by several researchers, like Sonquist [94], Messenger *et al.* [95], and Gillo [96, 97]. However, modern formulization and adaptations can be credited to the work of Breiman [98], Friedman [99], and Quinlan [100-102], who simplified and standardised the methodological framework of tree-based methods. This makes tree-based methods easier to comprehend and implement for researchers from diverse fields of interest.

All tree-based methods share the same principle of construction, which involves recursive partitioning of data or its associated characteristics based on some pre-defined criteria. To understand the basic formulation of a tree-based method,

consider a sample space $Z \in \mathcal{Z}$ that is composed of n features, that is, $z_0, z_1, z_2, \dots, z_n$. Its tree representation can be constructed by repeatedly splitting its features into smaller subsets based on some pre-defined criterion, as shown in Figure 3-1. This, provides an intuitive way to visualize the decision rules when applied to the analysis of categorical (classification tree) or continuous (regression tree) data.

Figure 3-1: Basic tree structure

The real success of trees-based methods is based on their inherent characteristics. They are non-parametric methods and can handle heterogeneous data. Therefore, they can be used to model any complex system with any variable type without considering any a priori assumptions. Tree-based methods can intrinsically implement feature selection and are robust to outliers or errors in labels. Thus, tree-based methods are robust to irrelevant or noisy variables and are easy to interpret, even by users with no statistical background.

3.1.2 Bootstrap aggregation

Bootstrap aggregation, also called bagging, is based on the concept of bootstrap, which was first introduced in 1979 by Bradley Efron [75] as an improvement to jackknife resampling methods [81, 103], as discussed in Chapter 2. As per the concept of bootstrap, the model generated from resampled data, which were previously sampled from an original dataset, follows the same characteristics as the main dataset.

This concept was further developed and used by researchers like Freeman *et al.* [104], Singh *et al.* [105], Efron *et al.* [106] and Breiman [107, 108]. Breiman in 1996 [107] suggested a method called bagging predictors, which uses the concept of bootstrap for predictor classification and then aggregates the resultant predictors. This combination of bootstrap and aggregation reduces the impact of high data variance and improves prediction accuracy.

Since then, the combination of aggregation and bootstrap has been used by many researchers and is commonly known as bootstrap aggregation or bagging.

3.1.3 Random forest

Random forest (RF) is a tree-based method that was initially introduced by Breiman *et al.* [108] as a classification technique. In his seminal paper, Breiman *et al.* [108] suggested combining the concept of bagging [75] with the random variable selection at each node. Thus, the resulting method (random forest) ensembles a large number of decision trees by averaging individual uncorrelated trees. This set of combinations results in reduced bias and high variances of individual decision trees.

In the same work [108], Breiman *et al.* demonstrated that random forests always converges as the number of trees in the forest goes to infinity using the law of large numbers. This, resolves the over-fitting problem that arises when using traditional tree-based techniques and allows its use for high-dimensional data.

Later, the concept of random forest was modified by Liaw and Wiener [109], who suggested the use of the random forest method for regression analysis. They proposed that instead of assigning class labels to tree predictor $T(\mathbf{x}; \boldsymbol{\theta})$ arising from growing trees based on random vector $\boldsymbol{\theta}$, as is done for classification, they should be assigned numerical values. Thus, it is assumed that the training sets are independently drawn from the distribution belonging to random vectors $Y, \mathbf{X}$.

Using this assumption, the mean-squared error for any numerical predictor $T(\mathbf{x})$ can be calculated using:

$$E_{\mathbf{X}, Y} = (Y - T(\mathbf{X}))^2$$

Moreover, the predictors are calculated by averaging $T(\mathbf{x}; \boldsymbol{\theta}_k)$ over a pre-selected number of k trees. Thus, it can be shown mathematically

$$\hat{f}_{rf}^B(\mathbf{x}) = \frac{1}{B} \sum_{k=1}^B T(\mathbf{x}; \boldsymbol{\theta}_k)$$

The basic implementation of the random forest method as a regularization technique as summarized by Hastie *et al.* [72] is as follows:

Pseudo-code to implement random forest

1. For a collection of tree-structured classifiers $T(\mathbf{x}; \boldsymbol{\theta}_k)$, for $k = 1, 2, 3, \dots, B$ where $\{\boldsymbol{\theta}_k\}$ are identically independent distributed vectors:
 - a. Bootstrap sample Z^* of size N from the training data.
 - b. Grown a random forest tree T_k to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
 - i. Select m variables at random from p variables.
 - ii. Pick the best variable/split-point among the m .
 - iii. Split the node into two child nodes.
 2. Output the ensemble of trees $\{T_k\}_{k=1}^B$
 3. For classification, let $\hat{C}_k(\mathbf{x})$ be the class prediction of the k^{th} random forest tree
- Then, $\hat{C}_{rf}^B(\mathbf{x}) = \text{majority vote } \{\hat{C}_k(\mathbf{x})\}_{k=1}^B$
-

Based on these findings, we extended the use of random forest as a regression technique for estimating the weight matrix W for Granger causality analysis. The overview of the proposed implementation is shown in Figure 3-2, whereas lasso, elastic-net, and standard OLS based Granger causality are implemented as discussed in Chapter 2.

Figure 3-2: Flow chart of Granger causality evaluation using random forest regularization

3.2 Performance Evaluation of Random Forest Granger Causality

3.2.1 Experimentation

We compared the effectiveness of our proposed method with existing lasso-regularized Granger causality and elastic-net regularized Granger causality methods. These methods were implemented in MATLAB, using multiple functions from different toolboxes. For random forest regularized Granger causality, we implemented the original formulation of the random forest method proposed by Breiman [109] on Python and MATLAB using the R package [108]. We then used it together with some functions from Granger causality connectivity analysis (GCCA) [68] for implementing Granger causality and followed the steps shown in Figure 3-2.

For implementing lasso Granger causality analysis, elastic-net Granger causality, and standard OLS Granger causality, we modified the GCCA toolbox. We

incorporated some regularization functions from the Sparse Learning with Efficient Projections (SLEP) toolbox [110].

3.2.2 Review of datasets

In order to evaluate the performance of our proposed method against existing methods, we used two simulated datasets previously used by other researchers to test similar kinds of algorithms [37, 69, 111, 112]. The use of these existing datasets allows us to remain unbiased in our comparison and does not raise questions about the reliability of the datasets.

We also applied the proposed method to two real human data sets, called StarPlus and HeLa cell datasets, in order to evaluate network structure involved in deductive reasoning and the cancer cell cycle respectively.

3.2.2.1 Simulated dataset

The first set of data was used by Mukhopadhyay *et al.* [113]; it simulates a scenario of fourteen variables and uses the following set of mathematical equations:

$$x_1(t) = 0.7 x_1(t - 1) + \varepsilon_1(t)$$

$$x_2(t) = 0.29 x_2(t - 1) + 0.65 x_1(t - 1) + \varepsilon_2(t)$$

$$x_3(t) = 0.15 x_3(t - 1) + 0.29 x_2(t - 1) + 0.65 x_{14}(t - 1) + \varepsilon_3(t)$$

$$x_4(t) = 0.17 x_4(t - 1) + 0.4 x_3(t - 1) + 0.7 x_6(t - 1) + \varepsilon_4(t)$$

$$x_5(t) = 0.6 x_5(t - 1) + 0.8 x_4(t - 1) + \varepsilon_5(t)$$

$$x_6(t) = 0.12 x_6(t - 1) + 0.3 x_7(t - 1) + 0.3 x_8(t - 1) + 0.3 x_9(t - 1) + \varepsilon_6(t)$$

$$x_7(t) = 0.8 x_7(t - 1) + \varepsilon_7(t)$$

$$x_8(t) = 0.7 x_8(t - 1) + \varepsilon_8(t)$$

$$x_9(t) = 0.77 x_9(t-1) + \varepsilon_9(t)$$

$$x_{10}(t) = 0.4 x_{10}(t-1) + 0.3 x_{11}(t-1) + \varepsilon_{10}(t)$$

$$x_{11}(t) = 0.7 x_{11}(t-1) + \varepsilon_{11}(t)$$

$$x_{12}(t) = 0.4 x_{12}(t-1) + 0.4 x_{11}(t-1) + \varepsilon_{12}(t)$$

$$x_{13}(t) = 0.4 x_{13}(t-1) + 0.4 x_{11}(t-1) + \varepsilon_{13}(t)$$

$$x_{14}(t) = 0.65 x_{14}(t-1) + \varepsilon_{14}(t)$$

The second set of simulated data was first used by Schelter *et al.* [112] and later in Seth [68], Siyal *et al.* [69], Tam *et al.* [114], and Baccala *et al.* [115]. It simulates a scenario of five variables and uses the following set of equations:

$$x_1(t) = 0.6 x_1(t-1) + 0.65 x_2(t-2) + \varepsilon_1(t)$$

$$x_2(t) = 0.5 x_2(t-1) - 0.3 x_2(t-2) - 0.3 x_3(t-4) + 0.6 x_4(t-1) + \varepsilon_2(t)$$

$$x_3(t) = 0.8 x_3(t-1) - 0.7 x_3(t-2) - 0.1 x_5(t-3) + \varepsilon_3(t)$$

$$x_4(t) = 0.5 x_4(t-1) + 0.9 x_3(t-2) + 0.4 x_5(t-2) + \varepsilon_4(t)$$

$$x_5(t) = 0.7 x_5(t-1) - 0.5 x_5(t-2) - 0.2 x_3(t-1) + \varepsilon_5(t)$$

As suggested by the above-mentioned authors, we generated the desired quantities of data from these sets of equations by first initializing the equations with white Gaussian noise having zero mean and unit variance. Then, we used the iterative process to generate the rest of the data.

Using the mathematical model discussed above, the direct and indirect influence structures of the simulated autoregressive models 1 and 2 are depicted by the directed graphs in Figure 3-3 and Figure 3-4, respectively. In both figures, a directed edge represents Granger causality between two nodes or variables.

Figure 3-3: Influence graph for simulated dataset 1.

Figure 3-4: Influence graph for simulated dataset 2.

3.2.2.2 Real dataset

HeLa cell line

HeLa cells consist of an eternalized, uninterruptedly cultured cell line of human cancer cells. In contrast to typical human cells, HeLa cells can survive infinitely in laboratory tissue cultures. This immortal characteristic gives them enormous significance in biomedical research. Since the latter half of the twentieth century, HeLa cells have been used all around the world in countless medical studies, including investigations to determine the nature of cancer, research and development of vaccines, mapping of genes, treatment of diseases, and examining the mechanisms involved in programmed cell death [116].

The human HeLa cell cycling dataset studied in this thesis is taken from an investigation by Whitfield *et al.* [117], who studied cell division in a human cancer cell line using gene transcriptional profiling. The main objectives of their study were to identify genes that are periodically expressed in the human cell cycle and to form a comprehensive directory of cell cycle genes that can be used as a starting point for other functional discoveries.

Using spotted cDNA microarrays containing 22,692 elements representing more than 16,000 different human genes or containing 43,198 elements representing more than 29,500 genes (estimated by UNIGENE clusters), Whitfield *et al.* [117] identified the list of periodically expressed genes and found that most of them had previously been reported as associated with the proliferative state of tumors. The experimental datasets of this study can be freely accessed from a public website [118].

In this thesis, we used their experiment 3 data-set, which has been used in the past by other researchers [49, 119]. The experiment 3 data-set identifies more than 1100 genes that are periodically expressed during the cancer cell cycle. As the data points were not collected at homogeneous intervals, they were interpolated by cubic smoothing splines [120] before being used, following researchers like Hlavácková-Schindler *et al.* [49] and Lozano *et al.* [119].

In this study, we used nineteen pre-selected genes from the HeLa cell data, which are a subset of genes regarded as highly influential and have been investigated by other researchers [49, 119, 121]. These nineteen preselected genes are PCNA, NPAT, E2F1, CCNE1, CDC25A, CDKN1A, BRCA1, DHFR, TYMS, CCNF, CCNA2, CDC20, STK15, BUB1B, CKS2, CDC25C, PLK, CCNB1, and CDC25B.

StarPlus Data

The real data set, StarPlus fMRI data, was collected by Marcel *et al.* [122] and is freely available [123]. This real dataset had been used by many researchers, such as Mitchell *et al.* [124], Wang *et al.* [125] and Hutchinson *et al.* [126]. It was developed to observe the workings of the brain during a task that involves understanding and comprehending the relationship between test sentences and pictures.

The study was performed on 13 normal subjects, and each individual was subjected to 40 trials. During each trial, each subject was shown a sentence that either correctly or incorrectly described the picture, and in a fixed time interval, he or she was asked to tell whether the sentence described the picture correctly or not. For the

first 20 trials, subjects were showed the sentence first, followed by the picture, and the order was reversed for the remaining 20 trials.

In each scenario, the first stimulus lasted for 4 seconds, followed by a blank screen for 4 seconds. Then, the second stimulus was given for 4 seconds, followed by resting time in which subject needed to decide the answer. Before the start of the next trial, each subject got a 15-second rest period, which resulted in around 30-seconds for each trial. Further details about experimental settings, sentences, and pictures can be found in Keller *et al.* [122].

During this experiment, T2-weighted fMRI images were collected at an interval of 500 msec using the 3T Signa scanner with TE =18ms and a flip angle of 50°. The resulting images captured approximately 5000 voxels per subject in eight oblique axial slices in two separate non-contiguous four-slice volumes. The first set was positioned to cover prefrontal areas and superior parietal, while the other set was positioned to cover the posterior temporal, inferior frontal, and occipital areas.

Before using the fMRI images for further analysis, the images were pre-processed to remove different artifacts (signal drift and head motion), which were evolved during the image-acquisition process using the FIASCO program [127]. After pre-processing, 25 anatomical regions of interest (ROIs) were selected, namely the left dorsolateral prefrontal cortex (LDLPFC) and right dorsolateral prefrontal cortex (RDLPFC), calcarine sulcus (CALC), left frontal eye fields (LFEF), right frontal eye fields (RFEF), left inferior parietal lobule (LIPL), right inferior parietal lobule (RIPL), left intraparietal sulcus (LIPS), right intraparietal sulcus (RIPS), left inferior frontal gyrus (LIFG), left opercularis (LOPER), right opercularis (ROPER),

supplementary motor areas (SMA), left and right inferior temporal lobule (LIT, RIT), left and right posterior precentral sulcus (LPPREC, RPPREC), left and right supramarginal gyrus (LSGA, RSGA), left temporal lobe (LT), right temporal lobe (RT), left and right triangularis (LTRIA, RTRIA), left superior parietal lobule (LSPL) and right superior parietal lobule (RSPL).

In this study, we restricted ourselves to seven ROIs for the fMRI StarPlus dataset, namely CALC, LDLPFC, LIPL, LIPS, LOPER, LT, and LTRIA, as these ROIs are considered most relevant by domain experts [125] in deductive reasoning.

3.3 Results and Discussion

For a systematic dataset, the performance of our proposed method against other existing methods is evaluated using the following measures: sensitivity (recall), specificity, precision, false positive rate (FPR), false discovery rate (FDR), accuracy, F1-Score and Matthews correlation coefficient (MCC). We used average values of 200 scenarios for comparison. These measures are calculated using

$$Specificity = \frac{TN}{TN+FP} ;$$

$$False\ Discovery\ Rate = \frac{FP}{TP+FP} ;$$

$$Precision = \frac{TP}{TP+FP} ;$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} ;$$

$$Recall = \frac{TP}{TP+FN} ;$$

$$F1 - Score = \frac{2 \times TP}{(2 \times TP) + FP + FN} ;$$

$$Matthews\ Correlation\ Coefficient = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} ;$$

where

TP = Number of true positive edges detected against ground truth network

TN = Number of true negative edges detected against ground truth network

FP = Number of false positive edges detected against ground truth network

FN = Number of false negative edges detected against ground truth network

For the simulated data set, a time series of length $T = 550$ was generated from the set of equations discussed in Chapter 3; by first initializing them using white Gaussian noise with zero mean and unit variance. The first 50 observations were excluded to remove the transient effect in data and the remaining 200-time points were used for further analysis.

For evaluating the best model order, we compared the AIC and BIC model validation methods. The reason for choosing AIC and BIC over other validation methods is because AIC and BIC utilize the whole dataset for evaluating model performance. In comparison, other methods partition data which could result in loss of data characteristics. To decide model selection between AIC and BIC, we evaluated AIC and BIC values for a range of model orders, using the following mathematical equations [68] :

$$AIC(p) = \ln(\det(\text{cov}(\text{residual error}))) + \frac{2pn^2}{T}$$

$$BIC(p) = \ln(\det(\text{cov}(\text{residual error}))) + \frac{\ln(T)pn^2}{T}$$

where n is the number of variables, p is the model order, and T is the number of observational time-points.

The results for the comparative study are shown in Figure 3-5 and Figure 3-6 for simulated datasets, Figure 3-7 for the HeLa cancer cell dataset, and Figure 3-8 for the real fMRI dataset. The findings suggest that the AIC yields better results compared to the BIC in terms of information loss and model complexity. These results are consistent with the findings of Yang [128], who suggested the use of AIC for selecting the model with the least mean squared error.

Once the model order was selected, the hyper-parameters for lasso and elastic-net regularization were tuned using default settings of the SLEP toolbox, which uses Armijo Goldstein's rule [129] for line search and the accelerated gradient method [130, 131], as mentioned in the SLEP toolbox manual [132].

For the random forest method, we performed 5-fold cross-validation by splitting data in 5 equal parts. In each iteration, 4 parts were used for training while the remaining part was used for testing. The model with the lowest error is selected for comparative study. This whole 5-fold cross validation process ensures that all data are used at least once for training. During the training process, hyper-parameters were tuned to $mtry = \text{floor}(\sqrt{\text{number of features}})$ and $ntree = 300$.

Figure 3-5: Information loss criterion plot using AIC and BIC for model selection for simulated dataset 1.

Figure 3-6: Information loss criterion plot using AIC and BIC for model selection for simulated dataset 2.

Figure 3-7: Information loss criterion plot using AIC and BIC for model selection for HeLa cell dataset

Figure 3-8: Information loss criterion plot using AIC and BIC for model selection for StarPlus dataset

The results for simulated dataset 1 are shown in Table 1 and Figure 3-9. The results of the comparative study using simulated data 2 are shown in Table 2 and Figure 3-10.

Table 1: Results of fourteen-variable scenario for comparative study of lasso Granger causality, elastic-net Granger causality, random forest Granger causality, and standard OLS Granger causality.

	<i>Lasso – based Granger Causality</i>	<i>Elastic-net – based Granger Causality</i>	<i>Random Forest – based Granger Causality</i>	<i>Standard OLS – based Granger Causality</i>
Accuracy	0.9562	0.9588	0.9705	0.9755
F1 Score	0.4388	0.4888	0.6923	0.7113
False discovery rate	0.0000	0.0000	0.0516	0.0205
Matthews correlation coefficient	0.5177	0.5564	0.7088	0.7392
Precision	1.0000	1.0000	0.9484	0.9795
Recall	0.2843	0.3269	0.5541	0.6101
Specificity	1.0000	1.0000	0.9977	0.9994

Table 2: Results of five-variable scenario for comparative study of lasso Granger causality, elastic-net Granger causality, random forest Granger causality, and Standard OLS Granger causality.

	<i>Lasso – based Granger Causality</i>	<i>Elastic-net – based Granger Causality</i>	<i>Random Forest – based Granger Causality</i>	<i>Standard OLS – based Granger Causality</i>
Accuracy	0.8592	0.9165	0.9227	0.9530
F1 Score	0.6184	0.7174	0.7772	0.9166
False discovery rate	0.3500	0.2167	0.2167	0.0667
Matthews correlation coefficient	0.5420	0.7095	0.7612	0.9030
Precision	0.6500	0.7833	0.7833	0.9333
Recall	0.6000	0.8000	0.8782	0.9784
Specificity	0.9000	0.9500	0.9500	1.0000

Figure 3-9: Results of the comparative study of lasso Granger causality, elastic-net Granger causality, random forest Granger causality, and standard OLS Granger causality on simulated dataset 1.

Figure 3-10: Results of comparative study of lasso Granger causality, elastic-net Granger causality, random forest Granger causality, and standard OLS Granger causality on simulated dataset 2.

For the real StarPlus fMRI dataset, the network structure detected using random forest Granger causality is shown in Figure 3-11 while the genetic map for Hela cell data is shown in Figure 3-12. The top 20 significant interactions are shown in Table 3.

Figure 3-11: Directed graph for real StarPlus dataset using seven ROIs using random forest Granger causality.

Figure 3-12: Directed map for nineteen pre-selected genes using random forest Granger causality

Table 3: Top 20 significant gene interactions using lasso regularized Granger causality, elastic-net regularized Granger causality and random forest Granger causality

<i>LASSO Regularized Granger Causality</i>	<i>Elastic-net Regularized Granger Causality</i>	<i>Random Forest Granger Causality</i>	<i>Standard OLS based Granger causality</i>
BRCA1↔PCNA	<i>CCNB1↔CDC25B</i>	PCNA↔CCNE1	PCNA↔BRCA1
E2F1↔NPAT	<i>PCNA↔BRCA1</i>	NPAT↔E2F1	NPAT↔BUB1B
PCNA↔TYMS	<i>CDC25B↔CDC25A</i>	E2F1↔NPAT	E2F1↔BRCA1
BRCA1↔CCNB1	<i>BRCA1↔CCNB1</i>	E2F1↔CCNE1	CCNE1↔BRCA1
CDC25B↔E2F1	<i>PCNA↔TYMS</i>	CCNE1↔CDC25A	CCNE1↔BUB1B
CKS2↔NPAT	<i>NPAT↔CDC25A</i>	CCNE1↔BUB1B	CCNF↔BUB1B
CDKN1A↔CDC25B	<i>CDC25B↔CDC25C</i>	CDC25A↔CCNB1	CCNA2↔BRCA1
BUB1B↔CDC25A	<i>DHFR↔CDC25A</i>	CDKN1A↔NPAT	CDC20↔BUB1B
CDC25B↔BUB1B	<i>BRCA1↔E2F1</i>	CDKN1A↔CCNE1	STK15↔BRCA1
E2F1↔BRCA1	<i>CDC25B↔PLK1</i>	BRCA1↔CCNE1	STK15↔BUB1B
CDC25B↔TYMS	<i>CDC25B↔TYMS</i>	BRCA1↔CCNB1	BUB1B↔BRCA1
NPAT↔CDC25B	<i>NPAT↔E2F1</i>	CCNF↔BUB1B	CKS2↔BRCA1
PLK1↔CDC25B	<i>CCNE1↔CKS2</i>	CCNA2↔NPAT	CDC25C↔BRCA1
BRCA1↔TYMS	<i>CDC25B↔BRCA1</i>	CDC20↔BUB1B	CDC25C↔BUB1B
CDC25B↔CCNF	<i>BUB1B↔CDC25B</i>	STK15↔CCNB1	PLK1↔BRCA1
DHFR↔TYMS	<i>CDC25B↔BUB1B</i>	CKS2↔CCNE1	PLK1↔BUB1B
DHFR↔CDC25A	<i>DHFR↔TYMS</i>	CKS2↔CDC25A	CCNB1↔BRCA1
CDC25A↔NPAT	<i>CCNF↔CDC25B</i>	CDC25C↔CCNE1	CCNB1↔BUB1B
CDC25B↔BRCA1	<i>CKS2↔NPAT</i>	TYMS↔BUB1B	TYMS↔BUB1B
CCNB1↔CDC25B	<i>STK15↔CDC25B</i>	DHFR↔CCNB1	DHFR↔BUB1B

Based on the results in Table 1 and Table 2, for five variable scenarios, our proposed method performs 2–15% better in terms of accuracy, FDR, F1-score, MCC, and recall when compared to lasso and elastic-net -based Granger causality. However, in terms of precision and specificity, random forest Granger causality and elastic-net -based Granger causality yield the same results. For the 14-variable system, the random forest based method performs 2–18% better in terms of accuracy, FDR, F1-score, MCC, and recall compared to lasso and elastic-net -based Granger causality. In terms of precision, lasso and elastic-net based methods outperform the random forest method and the results are approximately the same in terms of specificity. The probable reason for the better precision and specificity of lasso and elastic-net based methods could be the low recall rate.

Whereas, compared to standard OLS based Granger causality, the random forest based Granger causality analysis underperforms by 0.5–4% in all measures for both scenarios. However, this difference in performance can be ignored because of the versatility of the random forest based method, which allows its use for the analysis of both high- and low-dimensional data.

For the StarPlus fMRI data-set, there are two options to determine the plausibility of the connection between various ROIs as shown in Figure 3-11. The first option is to carry out a clinical trial to verify the proposed network map and the second option is to look for individual connections considering the known functionality of each region. As clinical trials were beyond the scope of this thesis, we determined plausibility based on the known functionality of each region.

The first ROI is the CALC, which houses the primary visual cortex. The primary function of the CALC is to obtain contrast of light from the retina and pass it to various regions of the brain for further processing. The second ROI is the LT, which is considered the speech and language recognition centre of the brain. Its functions include comprehension of speech and writing in terms of semantics, recognition of languages, numbers, and mathematical symbols.

The next ROI, LIPS, is located in the lateral parietal lobe of the brain [133] and is responsible for perceptual-motor coordination (e.g., directing eye movements and reading) and visual attention. The LOPER is part of Brodmann's area 44, which is responsible for interpreting reading languages [134], performing lexical decisions [135, 136], making semantic decisions [137], spelling [138], and differentiation between words and blanks [139].

The LIPL is connected to Broca's area and Wernicke's area, and acts as an intermediary between the two areas. It is responsible for classification and labelling of things and words, which is a pre-requisite for forming concepts and abstract thinking [140, 141]. The next ROI is the LDLPFC, which is involved in executive functions, like decision making [142], working memory management [143], planning [144], and cognitive processing [145]. The last ROI is the LTRIA, which is part of Broca's area. The LTRIA shows high activity during semantic decision tasks, like determining whether a word represents an abstract concept or is a concrete entity, and generation of tasks.

Using the functionality of the above-mentioned regions, we can infer high plausibility of the following links: $\text{CALC} \rightarrow \text{LIPS}$, $\text{LOPER} \rightarrow \text{LTRIA}$, $\text{LDLPFC} \rightarrow \text{LTRIA}$ and $\text{LOPER} \rightarrow \text{LIPL}$. However, based on functional knowledge of ROIs, other links between ROIs are likely to be false positive detections.

For the results on real HeLa data, it is not possible to compare the whole network; however based on the advice of domain experts and other researchers, we extracted the top 20 significant interactions detected by each method and cross-referenced them using the BioGRID database [146] to observe whether they had been reported earlier. The interactions that were found in the database are shown in bold in Table 3. Among these top 20 interactions, our proposed method detected 9 known interactions while lasso and elastic-net based methods detect 3 and 5 interactions, respectively. The OLS based method detected 10 known interactions.

The results of the performance evaluation of random forest Granger causality suggest that it is crucial to select the correct number of features to be analysed together with the correct number of trees used to assess Granger causality via random forest Granger causality. However, future researchers need to investigate this area in much greater detail to reveal the optimum relationship between number of features and number of trees.

3.4 Summary

In this chapter, we described our proposed method called random forest Granger causality, by discussing its formulation and implementation. We discussed the experimental details related to the performance evaluation of random forest Granger causality and the details related to simulated, StarPlus fMRI, and HeLa cell

line datasets, which are used for performance evaluation. We undertook a detailed comparison between existing and proposed methods, using simulated data and various statistical measures that include sensitivity (recall), specificity, precision, negative predictive value, FPR, FDR, accuracy, F1 – Score, and MCC. We used our proposed method to map the brain network structure involved in deductive reasoning structure and the gene network that is commonly found in cancer using real StarPlus data and real HeLa cell line data respectively. We discussed the results in detail to make an educated judgment about the effectiveness of the proposed method.

In the next chapter, we will discuss our second proposed method, called elastic-net copula Granger causality, which uses copula and elastic-net regularization to handle dimensionality and linearity issues related to the use of high-dimensional biological data.

4 COPULA GRANGER CAUSALITY

4.1 Introduction

Granger causality analysis was first introduced in econometrics [21] to study causal relationships among different financial time-series data. Classical literature, such as *New Introduction to Multiple Time Series Analysis* by Lütkepohl [26], showed that Granger causality can be applied using many modeling techniques, including VAR models, infinite order VAR, and impulsive response functions. However, the literature in bioinformatics indicates that the VAR modeling technique is commonly used because of its simplicity and ease of interpretation [54].

In the past, several methods have been used to handle high-dimensional data, such as pairwise analysis [114], kernel-based algorithms [147, 148], and other regularization-based methods [34, 84]; however they have not yielded the desired results. Another viable alternative, proposed by Nelsen [149], uses copula to discover dependencies between random variables. The use of probability theory in the form of copulas counters the spurious effect of confounders by utilizing marginal probabilities to incorporate prior information about them.

Recently, Liu *et al.* [150] introduced Gaussian copula with non-parametric marginals, which can be used to estimate high-dimensional undirected graphs for

inferring influences of high-dimensional time-series. This idea was developed by Bahadori and Liu [151], who proposed a new method for finding Granger causality, called the Gaussian non-paranormal (GNPN) model.

Bahadori and Liu [151] defined the GNPN model by considering a time series $X = (X_1, \dots, X_n)$ having a GNPN distribution of $GNPN(X, B, F)$ as long as there are functions $\{F_j\}_{j=1}^n$ such that $F_i(X_j)$ for $j = 1, \dots, n$ are jointly Gaussian and can be factored according to the VAR model through coefficients $B = \beta_{i,j}$. More precisely, variables $Z_i \triangleq F_i(X_j)$ can be factored as shown below:

$$p_z(z) = \mathcal{N}(z(1, \dots, L)) \times \prod_{j=1}^n \prod_{t=L+1}^T p_{\mathcal{N}}(z_j(t) | \sum_{i,j}^T \beta_{i,j}^T z_i^{t,Lagged}, \sigma_j)$$

where $p_{\mathcal{N}}(z|\mu, \sigma)$ is the Gaussian density function having variance σ^2 and mean μ , L is the maximal time delay, $z_i^{t,Lagged} = [z_i(t-L), \dots, z_i(t-1)]$ is the history of z_i until time t , and $\beta_{i,j} = [\beta_{i,j}(1), \dots, \beta_{i,j}(L)]$ is the vector of coefficients modelling the effect of time series z_i on the target time series.

Based on this information, the time series z_j Granger-causes z_i if at least one value in the coefficient vector β_j is statistically nonzero [151].

The idea behind using the GNPN model is to isolate the marginal properties of the data from their dependency structures. In order to implement this idea, Bahadori *et al.* [151] used ℓ_1 (lasso) regularization to estimate dependency structures for high-dimensional data. The authors used the advantage of lasso scalability for

higher dimensions and contemporaneously used copulas to handle non-linearity of the data.

4.2 Limitations of Existing Work

In the standard lasso estimator [89], the ℓ_1 penalty is used to obtain the sparse solution to the following optimization problem:

$$\beta(LASSO) = \underset{\beta}{\operatorname{argmin}} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

where $\|\beta\|_1 = \sum_{i=1}^p |\beta_i|$ is the ℓ_1 -norm penalty on β and induces sparsity in the solution, and $\lambda \geq 0$ is a tuning parameter.

The use of the ℓ_1 -norm penalty helps in the simultaneous operations of regularization and shrinkage, and thus, makes lasso an appealing variable selection method. However, despite these advantages, lasso faces limitations. According to Hui [92], Hui *et al.* [82], Fan *et al.* [152], and other researchers, the major limitations that make lasso unstable are as follows.

- 1) When the number of observational points (p) is greater than the number of variables (n), lasso limits variable selection to n variables before it becomes saturated [82].
- 2) When the number of variables (n) is greater than the number of observational points (p), lasso's prediction performance becomes dominated by ridge regression [82, 89].

- 3) When dealing with either highly correlated predictors or identical predictors, lasso usually results in random selection of predictors [153, 154].

4.3 Proposed Elastic-net Copula Granger Causality

The instability of lasso can be circumvented using an extension of lasso called elastic-net. It is robust to high correlations among predictors [92] and can select more than p variables when $n \gg p$ (*i.e.*, the number of variables $\gg$ the number of observation points). Elastic-net uses a mixture of ℓ_1 (lasso) and ℓ_2 (ridge regression) penalties, and can be formulated as

$$\beta(enet) = \underset{\beta}{\operatorname{argmin}} \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$$

where $\|\beta\|_1 = \sum_{i=1}^p |\beta_i|$, $\|\beta\|_2^2 = \sum_{i=1}^p \beta_i^2$, λ_1 is the tuning parameter for ℓ_1 (lasso), and λ_2 is the tuning parameter for ℓ_2 (ridge regression).

As we know, ℓ_2 (ridge regression) [91] works well with a large number of predictors that either have non-zero coefficients or are drawn from a normal distribution and are highly correlated [153]. Therefore, its presence in elastic-net helps improve variable selection, whereas the ℓ_1 (lasso) penalty induces the grouping effect and stabilizes solution paths with respect to random sampling [154]. Thus, combining both penalties should significantly improve predictions even when working with high-dimensional datasets.

Based on these findings, we propose a method called elastic-net copula Granger causality (ECGC), which we expect to be more precise than the existing

GNPN method. In this new method, we employ elastic-net regularization instead of ℓ_1 (lasso) regularization to estimate dependency structures. The use of elastic-net overcomes the shortcomings of existing lasso methods, and we exploit the advantages of copula to handle non-linearity of the data.

Therefore, instead of solving the optimization problem proposed by Bahadori and Liu [151], that is,

$$\min_{\beta} \sum_{t=L+1}^T \left\| X_i(t) - \sum_{i=1}^p \beta_{i,j}^T X_i^{t,Lagged} \right\|_2^2 + \lambda \|\beta\|_1$$

where λ is the tuning parameter, we propose using the following optimization problem:

$$\min_{\beta} \sum_{t=L+1}^T \left\| X_i(t) - \sum_{i=1}^p \beta_{i,j}^T X_i^{t,Lagged} \right\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2$$

where λ_1 and λ_2 are tuning parameters for ℓ_1 (lasso) and ℓ_2 (ridge regression) penalties, respectively.

4.4 Proposed Random Forest Copula Granger Causality

Based on findings discussed in the last chapter, we found that use of random forest as a regularization technique for evaluating Granger causality yields better results compared to the elastic-net based method. Therefore, we propose another variant of a copula based Granger causality method called random forest copula Granger causality.

The pseudo-code for implementing both variants of copula Granger causality is as follows.

Pseudo-code for implementing variants of copula Granger causality

1. Find the marginal distribution for each time series:

$$\hat{F}_n(t) = \frac{\text{number of elements in the samples} \leq t}{n}$$

2. Map the observations into copula space:

$$\hat{f}_i(\mathbf{X}_i^t) = \hat{\mu}_i + \hat{\sigma}_i \Phi^{-1}(\hat{F}_i(\mathbf{X}_i^t))$$

In practice, as proposed by Bahadori and Liu [151], to avoid large numbers $\Phi^{-1}(\mathbf{0}^+)$ and $\Phi^{-1}(\mathbf{1}^1)$; we use the winsorized estimator of the distribution function:

$$\tilde{F}_j = \begin{cases} \delta_n, & \text{if } \hat{F}_j(\mathbf{X}_j) < \delta_n \\ \hat{F}_j(\mathbf{X}_j), & \text{if } \delta_n \leq \hat{F}_j(\mathbf{X}_j) \leq 1 - \delta_n \\ (1 - \delta_n), & \text{if } \hat{F}_j(\mathbf{X}_j) > 1 - \delta_n \end{cases}$$

3. For elastic-net variant: evaluate Granger causality using elastic-net regularization and copula for different values of tuning parameters among $\hat{F}_i(\mathbf{X}_i^t)$.

- a. Use the cyclical coordinate descent method for parameter selections of elastic-net regularization.

For random forest variant: evaluate Granger causality using random forest regularization and $\hat{F}_i(\mathbf{X}_i^t)$.

4. Select Granger causality based on the minimum AIC [73].
-

4.5 Performance of Proposed Variants of Copula Granger Causality

4.5.1 Experimentation

For implementation of our proposed method, we used MATLAB together with the well-known SLEP toolbox [110], which has functions related to regularization and has been used in the past by many researchers [155, 156]. Lasso copula Granger causality is implemented with the help of existing code found in Liu *et al.* [157] together with the Glmnet toolbox [158]. Standard OLS based Granger causality is implemented as discussed in Chapter 2 and random forest Granger causality is implemented as discussed in Chapter 3.

In order to select the model order for Granger causality analysis, we performed a comparative study between AIC and BIC, like that discussed in Chapter 3. Based on the results shown in Figure 4-1 to Figure 4-4, we used the AIC for model selection. The fine-tuning of hyper-parameters was performed using the cyclical coordinate descent method, as discussed in Chapter 3.

In order to remain unbiased in our comparison, we used data sets previously used by other researchers to test similar kinds of algorithms. We applied our method to a real HeLa cell data set and the StarPlus data-sets. Details related to the datasets can be found in Chapter 3 of this thesis.

4.6 Results and Discussion

Performance evaluation of both proposed methods was undertaken by comparing our proposed methods with the existing method using the following measures: precision, FDR, recall, and F1-score. To minimize unexpected errors, these measures were calculated for 1000 resampled data, and then their average values were used as results. These measures are calculated using the formulae discussed in Chapter 3.

The results for simulated dataset 1 are shown in Table 4 and Figure 4-5, while, the results for simulated dataset 2 are shown in Table 5 and Figure 4-6. The results for the StarPlus dataset, which maps the effective brain network involved in deductive reasoning, are presented in Figure 4-7 and Figure 4-8.

For the real HeLa cell data, we extracted the top 20 significant interactions. We looked up those 20 interactions in the BioGRID database [146] to observe whether they had been reported earlier. Once all 20 interactions were analysed, we compared the number of matches found by each method. This approach of evaluation has been used and suggested by other researchers [159]. The results of our proposed methods, the lasso copula method and standard OLS based method, are summarized in Table 6.

Figure 4-1: Information loss vs model order plot using AIC and BIC for model selection for simulated dataset 2.

Figure 4-2: Information loss vs model order plot using AIC and BIC for model selection for simulated dataset 1.

Figure 4-3: Information loss vs model order plot using AIC and BIC for model selection for HeLa cancer cell dataset.

Figure 4-4: Information loss vs model order plot using AIC and BIC for model selection for fMRI StarPlus dataset.

Table 4: Results for comparative study of elastic-net copula Granger causality, lasso copula Granger causality, random forest copula Granger causality, and standard OLS copula Granger causality on simulated data set 1.

	<i>Lasso – based Granger Causality</i>	<i>Elastic-net – based Granger Causality</i>	<i>Random Forest – based Granger Causality</i>	<i>Standard OLS – based Granger Causality</i>
Accuracy	0.9568	0.9573	0.9712	0.9755
F1 Score	0.4514	0.4591	0.7023	0.7086
False discovery rate	0.0007	0.0000	0.0587	0.0244
Matthews correlation coefficient	0.5274	0.5332	0.7160	0.7365
Precision	0.9993	1.0000	0.9413	0.9756
Recall	0.2949	0.3020	0.5693	0.6094
Specificity	1.0000	1.0000	0.9974	0.9993

Table 5: Results for comparative study of elastic-net copula Granger causality, lasso copula Granger causality, random forest copula Granger causality and standard OLS copula Granger causality on simulated data set 2.

	<i>Lasso – based Granger Causality</i>	<i>Elastic-net – based Granger Causality</i>	<i>Random Forest – based Granger Causality</i>	<i>Standard OLS – based Granger Causality</i>
Accuracy	0.8865	0.9253	0.962	0.9766
F1 Score	0.6792	0.7498	0.9004	0.9294
False discovery rate	0.1766	0.0100	0.0657	0.0161
Matthews correlation coefficient	0.6303	0.7478	0.8834	0.924
Precision	0.8234	0.9900	0.9343	0.9839
Recall	0.5910	0.6360	0.8839	0.8998
Specificity	0.9604	0.9976	0.9815	0.9958

Figure 4-5: Results for comparative study of elastic-net copula Granger causality, lasso copula Granger causality, random forest copula Granger causality, and standard OLS copula Granger causality on simulated data set 1.

Figure 4-6: Results for comparative study of elastic-net copula Granger causality, lasso copula Granger causality, random forest copula Granger causality, and standard OLS copula Granger causality on simulated data set 2.

Figure 4-7: Brain network involved in deductive reasoning using elastic-net copula Granger causality.

Figure 4-8: Brain network involved in deductive reasoning using random forest copula Granger causality.

Table 6 : Top 20 Significant gene interactions using lasso copula Granger causality, elastic-net copula Granger causality, random forest copula Granger causality, and standard OLS based copula Granger causality.

<i>Lasso Copula Granger Causality</i>	<i>Elastic-net Copula Granger Causality</i>	<i>Random Forest Copula Granger Causality</i>	<i>Standard OLS based Copula Granger Causality</i>
CCNB1↔CDC25B	CCNB1↔CDC25B	PCNA↔CCNA2	NPAT↔E2F1
E2F1↔CCNE1	E2F1↔CCNE1	NPAT↔CCNA2	CCNE1↔NPAT
CCNE1↔CDC25A	CCNE1↔CDC25A	E2F1↔CCNA2	TYMS↔CDC25B
PLK1↔CCNB1	PLK1↔CCNB1	E2F1↔BUB1B	CCNE1↔CKS2
PCNA↔NPAT	CDKN1A↔BRCA1	CCNE1↔CCNB1	CDC25A↔BUB1B
PCNA↔PCNA	PCNA↔NPAT	CDC25A↔CCNA2	CDKN1A↔TYMS
CDC25A↔CDKN1A	CDC25A↔CDKN1A	CDKN1A↔CCNA2	BRCA1↔E2F1
CDKN1A↔BRCA1	PCNA↔PCNA	BRCA1↔CCNA2	BRCA1↔CKS2
BRCA1↔CDC25B	CCNB1↔CCNF	CCNF↔CCNA2	CCNA2↔CDC20
CDC25C↔PLK1	CDC25C↔PLK1	CCNA2↔CCNE1	CCNA2↔CDC25C
CDC25B↔TYMS	CDC25B↔TYMS	CDC20↔CCNA2	CCNE1↔CDC25A
NPAT↔E2F1	CCNB1↔STK15	STK15↔CDKN1A	CDC20↔CCNA2
BUB1B↔CKS2	BUB1B↔CKS2	STK15↔CCNB1	CDC20↔CDC25C
DHFR↔DHFR	NPAT↔E2F1	BUB1B↔NPAT	CKS2↔CDC25B
CDC20↔CDC25B	DHFR↔DHFR	CKS2↔CDKN1A	CDC25C↔PCNA
CCNA2↔CDC20	STK15↔BUB1B	CCNB1↔CCNA2	CDC25C↔CCNE1
STK15↔BUB1B	CCNA2↔CDC20	CDC25B↔BRCA1	PLK1↔CCNB1
CKS2↔CDC25C	CKS2↔CDC25C	CDC25B↔STK15	CDC25B↔CCNF
NPAT↔NPAT	PCNA↔E2F1	TYMS↔CCNE1	TYMS↔CCNB1
BUB1B↔CDC25B	CCNB1↔CKS2	TYMS↔CCNA2	CDC20↔CCNF

Based on the results shown in Table 4 and Table 5, both of our proposed methods outperform the existing lasso based method by 3–20% in terms of accuracy, recall, F1-score, MCC, specificity, FDR, and precision. When comparing the performance of both newly proposed methods, elastic-net copula Granger causality and random forest Granger causality, with existing lasso copula Granger causality, we found that the random forest based method performs approximately 1–4 % better in terms of accuracy, 2.5% better in terms of recall, 1.5–2.5% better in terms of F1-score, and 13–18 % better in terms of MCC. At the same time, when comparing the two proposed methods, we observed that the elastic-net based method performs better than the random forest method by approximately 1% in terms of specificity, and 5.5% in terms of precision, and FDR. The better performance of elastic-net copula Granger causality can be attributed to a lower recall rate, which is consistent with the findings of chapter 3.

When comparing the results with standard OLS copula Granger causality, we found all regularization based methods underperformed in all measures except FDR, precision, and specificity. In the case of FDR and precision, the elastic-net based method outperforms OLS by approximately 0.5–2% and 0.5–2.5% respectively. In terms of specificity, OLS based variants have approximately the same results as elastic-net copula Granger causality. However, in terms of other measures, we found that random forest copula Granger causality yields comparative results and can be a better alternative to other regularization based methods, especially when dealing with high-dimensional data.

For the results on real HeLa data (among those top 20 significant interactions listed in Table 6), the elastic-net method detected seven previously reported interactions, whereas the lasso-based method found only five. These matched interactions are shown in bold in Table 6. Of the seven interactions found by the elastic-net method, we observed two interesting interactions (highlighted in Table 6) that were not detected at all by the lasso copula method. These interactions are related to different cancer cell cycles, as reported in Wolthuis *et al.* [160], Kong *et al.* [161], and Radulovic *et al.* [162], whereas other detected interactions were not found in the BioGRID database. For random forest, we were able to detect 5 known interactions whereas 7 known interactions were detected using the OLS based method.

For StarPlus fMRI dataset, based on the functional knowledge discussed in chapter 3, the plausible interactions among ROIs are $CALC \leftrightarrow LT$, $LOPER \rightarrow LT$, and $LOPER \rightarrow LDLPFC$ for the brain network detected by elastic-net copula Granger causality and $CALC \rightarrow LTRIA$, $LOPER \rightarrow LDLPFC$, $LIPS \rightarrow LOPER$, and $CALC \rightarrow LIPS$ for the brain network detected by random forest Granger causality.

4.7 Summary

In this chapter, we reviewed the use of copula for evaluating Granger causality. We discussed the existing lasso-based copula Granger causality implementation and some of its limitations. In order to overcome those limitations, we proposed two improved copula-based methods called elastic-net copula Granger causality and random forest copula Granger causality.

We performed a detailed comparative study between existing and proposed methods using simulated datasets with the following measures: sensitivity (recall), precision, FDR, and F1-score. Based on the results of these simulated studies, we can safely assert that our proposed methods outperform existing regularization based methods in all measures. Among the two proposed methods, the random forest Granger causality method performs better than the elastic-net copula method. We used both proposed methods to analysis real biological networks that include StarPlus fMRI data for mapping brain networks involved in deductive reasoning and HeLa cell line data for mapping gene networks involved in cancer.

In the next chapter, we will discuss the third contribution of this thesis called bi-directional Granger causality and its variants.

5 BI-DIRECTIONAL

GRANGER CAUSALITY

5.1 Introduction

High-throughput technologies, such as DNA microarrays and fMRI techniques, yield high-dimensional data by simultaneously analysing multiple variables. This makes use of the standard OLS implementation of Granger causality analysis unfeasible, because there are more variables than there are observational time points. In order to resolve this problem, alternative methods have been proposed, such as pairwise analysis [114], and regularization [49, 50], as previously discussed. In addition, Cheng *et al.* [159] proposed a method to handle the data shortage problem, called naïve forward-backward Granger causality, based on the idea that time-series can be reused by reversing the time stamp of the time-series, and more causal information can be extracted. Similar alternatives were discussed by Haufe *et al.* [163] and Hu *et al.* [164].

5.2 Naïve Forward-Backward Granger Causality

In 2014, Cheng *et al.* [159] suggested using the concept of Granger causality in a slightly different manner. Instead of looking at the past, they proposed considering future impacts of current actions. This results in extraction of more causal

information from time-series data. The authors called this method naïve forward backward Granger causality (shown in Figure 5-1).

Figure 5-1: Depiction of the concept behind bi-directional Granger causality

In explaining implementation of their proposed method, Cheng *et al.* [159] used the assumption that the original time series meets all necessary conditions to perform Granger casualty analysis, as studied in Bahadori and Liu [151] and Eichler [165], including linearity and stationarity of the time-series data. Once all conditions are validated, the authors proposed using the pseudo-code shown below, which uses the lasso-based Granger causality analysis algorithm [166].

Pseudo code for implementing naïve forward backward Granger causality.

INPUT :	Time series $\{y^t\}$, Lag P , Penalty parameter λ .
OUTPUT :	Coefficient L_i^{FB} , $l = 1, 2, \dots, P$.
METHODOLOGY :	<p>Define the backward time series $\{z^t\}$ by $z^t = y^{-t}$.</p> <p>Get forward coefficient A_l via LASSO - Granger Causality with $\{y^t\}$, P, and λ.</p> <p>Get backward coefficient B_l via LASSO - Granger Causality with $\{z^t\}$, P, and λ.</p> <p>Return $L_i^{FB} = \frac{1}{2} (A_l + B_l^T)$, $l = 1, 2, 3, \dots, P$.</p>

However, as discussed in Chapters 3 and 4, the use of an ℓ_1 -penalty function for evaluating Granger causality faces limitations when dealing with highly correlated data and variable selection. Therefore, we proposed two more efficient alternatives, namely

- 1 forward backward pairwise Granger causality, and
- 2 bi-directional random forest Granger causality.

5.3 Forward Backward Pairwise Granger Causality

5.3.1 Pairwise Granger causality

Pairwise Granger causality is an extension of standard OLS-based Granger causality analysis, as discussed in Chapter 2, and was suggested by Tam *et al.* [51, 114]. They proposed that for high-dimensional data, we can analyse variables in pairs and combine all the resultant matrices to form a single matrix, thereby by-passing the OLS limitation of handling high-dimensional data (*i.e.*, when the number of variables is greater than the number of observational time points).

In order to determine effectiveness of different variants of Granger causality analysis, we performed a detailed comparative study comparing lasso Granger causality analysis with other commonly used alternatives, including elastic-net Granger causality, pairwise Granger causality, and standard OLS-based Granger causality [69]. In this study, we compared these four methods using different statistical measures, namely, precision, FDR, F1-Score, and accuracy. These measures were calculated using the mathematical equations discussed in Chapter 3. We used multiple simulated datasets with five and fourteen variable networks, and

observed the effect of the VAR model order for each method. The findings of this experiment are shown in Figure 5-2, and Figure 5-3.

Figure 5-2: Precision and false discovery rate results for a comparative study among lasso, elastic-net, pairwise, and standard OLS – based Granger causality.

Figure 5-3: Accuracy and F1-score results for a comparative study among lasso, elastic-net, pairwise, and standard OLS – based Granger causality.

Based on the results of the comparative study shown above, we observed that a second-order VAR model yields better results. In terms of methods, OLS-based pairwise Granger causality yields results that are approximately close to standard OLS-based Granger causality. Therefore, based on this conclusion, we propose our first variant of bi-directional Granger causality, called forward backward pairwise Granger causality (FBPGC) [167], which is scalable to large datasets because of the use of pairwise analysis. The pseudo-code for implementing forward backward pairwise Granger causality is as follows.

Pseudo code for implementing forward backward pairwise Granger causality

INPUT :	Time series $\{y^t\}$, Lag P , Penalty parameter λ .
OUTPUT :	Coefficient PW_l^{FB} , $l = 1, 2, \dots, P$.
METHODOLOGY :	<p>Define the backward time series $\{z^t\}$ by $z^t = y^{-t}$.</p> <p>Get forward coefficient A_l via <i>Pairwise Granger Causality</i> with $\{y^t\}$, P, and λ.</p> <p>Get backward coefficient B_l via <i>Pairwise Granger Causality</i> with $\{z^t\}$, P, and λ.</p> <p>Return $PW_l^{FB} = \frac{1}{2} (A_l + B_l^T)$, $l = 1, 2, 3, \dots, P$.</p>

5.4 Bi-directional Random Forest Granger Causality

Our second proposed variant of bi-directional Granger causality is called bi-directional random forest Granger causality [168]. It combines re-utilisation of time-series data in conjunction with random forest Granger causality, as suggested by Furqan and Siyal [11] and discussed in Chapter 3. As this method uses random forest as a regularization method, it can be used to analyse many time-series. The

implementation of bi-directional random forest Granger causality uses the following pseudo-code.

Pseudo code for implementing bi-directional random forest Granger causality

INPUT :	Time series $\{y^t\}$, Lag P , Penalty parameter λ .
OUTPUT :	Coefficient RF_l^{FB} , $l = 1, 2, \dots, P$.
METHODOLOGY :	<p>Define the backward time series $\{z^t\}$ by $z^t = y^{-t}$.</p> <p>Get forward coefficient A_l via Random Forest Granger Causality with $\{y^t\}$, P, and λ.</p> <p>Get backward coefficient B_l via Random Forest Granger Causality with $\{z^t\}$, P, and λ.</p> <p>Return $RF_l^{FB} = \frac{1}{2} (A_l + B_l^T)$, $l = 1, 2, 3, \dots, P$.</p>

5.5 Performance Evaluation

We implemented the proposed methods and existing techniques in MATLAB using functions from the Granger causal connectivity analysis GCCA toolbox [68] and Glmnet toolbox [158]. We compared these methods in terms of precision and FDR against ground truth networks. The methods used in this comparative study include lasso-based naïve forward backward Granger causality (FBGC), pairwise forward backward Granger Causality (PFBGC), bi-directional random forest Granger causality (BRFGC) and standard OLS-based forward backward Granger causality (OLS-based FBGC).

5.5.1 Simulated dataset

For each study, we simulated 250 datasets of variable lengths ranging from 50-time points to 500-time points, using the simulated systems discussed in Chapter 2. Each time, the system was initialized with white Gaussian noise having unit variance and zero mean. The first 50-time points were then truncated to reduce the transient effect of the data.

5.5.2 Real biological dataset

We applied our proposed techniques to both real data sets discussed in Chapter 2, *i.e.* brain and gene datasets. For the gene dataset, we analysed 19 pre-selected genes as suggested by Hlavácková-Schindler and Bouzari [49], Hlavácková-Schindler and Pereverzyev [88], and Li et al [121]: PCNA, NPAT, E2F1, CCNE1, CDC25A, CDKN1A, BRCA1, CCNF, CCNA2, CDC20, STK15, BUB1B, CKS2, CDC25C, PLK1, CCNB1, CDC25B, TYMS, and DHFR. For brain effective connectivity mapping, we restricted our study to seven ROIs that were used and advised by other researchers [11, 125]. These seven ROIs are LIPL, LDLPFC, CALC, LTRIA, LT, LOPER, and LIPS.

5.6 Results and Discussions

5.6.1 Simulated dataset

Based on the simulated study comparing forward backward pairwise Granger causality with lasso Granger causality, we observed that our proposed method

showed more than 15% better precision and around 15% less FDR. These results are shown in Figure 5-4.

Figure 5-4: Results of a comparative study between forward backward pairwise Granger causality and lasso-based forward backward Granger causality.

The results of the comparative study between bi-directional random forest Granger causality and lasso-based forward backward Granger causality are shown in Figure 5-5. According to these results, lasso-based forward backward Granger causality on average yields approximately 38% precision, and 62% FDR, compared to our proposed method which yields 51% precision, and 49% FDR. Based on these results, our proposed bi-directional random forest Granger causality (RFGC) outperforms the existing lasso-based method.

The results of the comparative study between forward backward standard OLS-based Granger causality and naïve forward-backward Granger causality are depicted in Figure 5-6. These results show that standard OLS-based forward backward Granger causality outperforms lasso-based forward backward Granger causality with approximately 10.03% better precision and FDR.

Figure 5-5: Results of a comparative study of bi-directional random forest Granger causality vs. lasso-based forward backward Granger causality.

Figure 5-6: Results of a comparative study of standard OLS-based forward backward Granger causality vs. lasso-based forward backward Granger causality.

Based on the combined results shown in Figure 5-7, both of our proposed variants outperformed the original method. However, all of our proposed variants outperformed the original lasso-based method. The best performing method proves to be the standard OLS-based method, followed by pairwise forward backward

Granger causality, bi-directional random forest Granger causality, and least performing is lasso-based naïve forward backward Granger causality.

Figure 5-7: Combined results of simulated study on bi-directional Granger causality

5.6.2 Real dataset

The results map for gene networks using forward backward pairwise Granger causality and bi-directional Granger causality are shown in Figure 5-8 and Figure 5-9, respectively.

Figure 5-8: Inferred gene network for 19 preselected genes using forward backward pairwise Granger causality.

Figure 5-9: Gene network map found using bi-directional random forest Granger causality.

The effective connectivity map using forward backward pairwise Granger causality and bi-directional Granger causality are shown in Figure 5-10 and Figure 5-11, respectively.

Figure 5-10: Effective brain connectivity map using forward backward pairwise Granger causality.

Figure 5-11: Effective brain connectivity map using bi-directional random forest Granger causality.

Based on functional knowledge of ROIs as discussed in Chapter 3, our resultant networks in Figure 5-10 and Figure 5-11 show the connection between CALC and LIPS. This connection seems to transfer visual information (picture or sentence displayed on screen). The link between LOPER and LIPS is responsible for recognizing objects and words. The connection between Brodmann areas 44 and 45 shows movement of information from area 44 to area 45 for further information processing.

Other links, such as those from Brodmann area 45, represent transfer of information to LDLPFC, LIPL, and LT for further processing to evaluate meaning, relation and deduction of the task performed. The remaining connections exchange information related to finger movements for registering the answer to the task.

5.7 Summary

In this chapter, we discussed the standard concept of Granger causality followed by reinterpretation of the Granger causality concept by Cheng *et al.* [159]. We proposed two heuristic variants of bi-directional Granger causality called forward backward pairwise Granger causality and bi-directional random forest Granger causality which help deal with large-dimensional data by the reuse of time-series and the better extraction of causal information. We demonstrated the effectiveness of our proposed methods over existing methods using simulated data-sets, a real StarPlus dataset to map the network involved in deductive reasoning, and a real HeLa cell line dataset to map the gene network commonly found in different cancers.

In the next chapter, we conclude the thesis by summarizing the findings of our research and discussing areas where this work can be extended.

6 CONCLUSION AND FUTURE WORK

6.1 Conclusion

Network inference is one of the most dynamic research areas in the field of bioinformatics. It is used to study different biological systems, including gene expression, protein syntheses, metabolic formations, and neuronal communications. The advancement in technology yields multidimensional data because new techniques can simultaneously analyse multiple variables. Mapping these variable networks provides understanding of the systematic behaviors of disease susceptibility and possible treatments.

In order to study these networks, there are numerous ways to analyse the causal relationships among the constituent elements. One of the most common and widely-used principles is Granger causality analysis, which uses time-series data to analyse causal links between variables and maps the network using graph theory. However, with all its advantages, Granger causality faces limitations when handling high-dimensional and non-linear data.

In order to minimize these limitations, we introduced three heuristic methods, as discussed earlier in this thesis. The summary of each contribution is as follows.

- Random forest Granger Causality is the first major contribution of the thesis, and is presented in Chapter 3. This technique is capable of effectively handling both high- and low- dimensional data using random forests as a regularization technique for evaluating Granger causality. Based on our experiments, we believe that the effectiveness of random forest Granger causality is due to the robustness of random forests, as it is invariant to attribute scales (units) and insensitive to outliers. These properties stem from the natural ability of regularized trees to handle numerical features and data interactions. However, we found that random forest Granger causality is highly sensitive to correct selection of hyper parameters (number of features and number of trees). Even slight changes in hyper-parameter selection results in high variation in the predicted model.
- Elastic-net copula Granger causality and random forest copula Granger causality comprise the second major contribution, and are presented in Chapter 4. Both techniques are capable of handling both high- and low-dimensional linear data effectively. Elastic-net copula Granger causality uses elastic-net regularization for handling data dimensionality, whereas random forest copula Granger causality uses random forest regularization. However, copula handles linearity issues of data by normalizing the effect of individual distributions of variables.

- As our third and final contribution, in Chapter 5, we exploited the notion suggested by Cheng *et al.* [103] that, time-series data can be recycled by reversing the time stamps to study Granger causality. We used this notion together with existing and proposed Granger causality methods. We performed a bidirectional study using pairwise Granger causality and random forest Granger causality. We compared their effectiveness against an existing lasso-based method and found that the bi-directional pairwise Granger causality out-performed other variants. The reason for this outcome is because forward backwards pair-wise Granger causality uses OLS instead of regularization methods which use unconstrained minimization of least-squares. However, apart from inherent limitations of individual variants of Granger causality analysis, all bi-directional methods require evaluating Granger causality twice, thus making them computationally expensive.

6.2 Future Work

- During our study regarding use of random forests for evaluating Granger causality, we observed that random forest regularization is very sensitive to number of features and number of trees. We used the default values prescribed by Liaw and Wiener [109], but we believe that further investigation is required to obtain the optimal relationship between number of features and number of trees used during the regularization process.
- We can extend the application of these algorithms in the area of social networking, where we can use them for predicting potential friends or related applications and advertisements.

- This work can also be extended in the field of econometrics where it can be used to infer the relationship of various aspects of an economy, such as to study the factors involved in deciding housing prices or the effect of inflation on housing prices and housing-related industries.

LIST OF PUBLICATIONS

Journal paper

- I. M. S. Furqan, and M. Y. Siyal, "Random forest Granger causality for detection of effective brain connectivity using high-dimensional data," *Journal of Integrative Neuroscience*, vol. 15, no. 01, pp. 55-66, 2016.
- II. M. S. Furqan, and M. Y. Siyal, "Inference of biological networks using bi-directional random forest Granger causality," *SpringerPlus*, vol. 5, no. 1, pp. 514, 2016.
- III. M. S. Furqan and M. Y. Siyal, "Biological network detection using forward backward pair-wise Granger causality," *International Journal of Simulation-Systems, Science and Technology*, vol. 17, 2016.
- IV. M. S. Furqan, and M. Y. Siyal, "Elastic-net copula Granger causality for inference of biological networks," *PLOS ONE*, vol. 11, no. 10, pp. e0165612, 2016.
- V. M. S. Furqan and M. Y. Siyal, "Inference of gene network using decision tree based Granger causality," *Molecular and Cellular Biomechanics*, (Under Review)
- VI. M. S. Furqan and M. Y. Siyal, " Mapping of effective brain connectivity using a copula-based Granger causality", *Journal of Integrative Neuroscience*, (First Review Completed)

- VII. M. S. Furqan and M. Y. Siyal, " Comparative analysis of variants of bi-directional Granger causality used in bioinformatics," *SpringerPlus*, (*Under Review*)

Conference papers

- I. M. S. Furqan and M. Y. Siyal, "Protein Map of Control Mice Exposed to Context Fear Using a Novel Implementation of Granger causality," in IEEE 3th International Conference on Artificial Intelligence, Modelling and Simulation, Sabah, Malaysia, 2015
- II. M. S. Furqan and M. Y. Siyal, "Gene Network Inference Using Forward Backward Pairwise Granger causality," in IEEE 3th International Conference on Artificial Intelligence, Modelling and Simulation, Sabah, Malaysia, 2015.
- III. M. Y. Siyal, M. S. Furqan, and S. M. G. Monir, "Granger causality: Comparative analysis of implementations for Gene Regulatory Networks," in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, 2014, pp. 793-798.

REFERENCES

- [1] S. R. Proulx, D. E. L. Promislow, and P. C. Phillips, "Network thinking in ecology and evolution," *Trends in Ecology & Evolution*, vol. 20, pp. 345-353, 2005.
- [2] G. Buzsaki, *Rhythms of the brain*: Oxford University Press, 2006.
- [3] B. Gibb and R. Guides, *The rough guide to the brain*: Rough Guides, 2007.
- [4] O. Sporns, "Networks analysis, complexity, and brain function," *Complexity - Special issue: Selection, tinkering, and emergence in complex networks*, vol. 8, pp. 56-60, 2002.
- [5] D. R. Chialvo, "Critical brain networks," *Physica A: Statistical Mechanics and its Applications*, vol. 340, pp. 756-765, 2004.
- [6] D. R. Chialvo, P. Balenzuela, and D. Fraiman, "The brain: what is critical about it?," *arXiv preprint arXiv:0804.0032*, 2008.
- [7] M. P. v. d. Heuvel and O. Sporns, "Rich-club organization of the human connectome," *The Journal of Neuroscience*, vol. 31, pp. 15775-15786, 2011.
- [8] B. Horwitz, "The elusive concept of brain connectivity," *NeuroImage*, vol. 19, pp. 466-470, 2003.
- [9] K. J. Friston, "Functional and effective connectivity: a review," *Brain Connectivity*, vol. 1, pp. 13-36, 2011.

- [10] S. I. f. Neurotechnology. (2016). *Cognitive engineering : Multi-modal connectivity brain network*. Available: <http://sinapseinstitute.org/projects/cognitiveengr/connectivity.php>
- [11] M. S. Furqan and M. Y. Siyal, "Random forest Granger causality for detection of effective brain connectivity using high-dimensional data," *Journal of Integrative Neuroscience*, vol. 15, pp. 55-66, 2016.
- [12] K. J. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Human Brain Mapping*, vol. 2, pp. 56-78, 1994.
- [13] R. A. Harvey and D. R. Ferrier, *Biochemistry*: Lippincott Williams & Wilkins, 2011.
- [14] M. Hecker, S. Lambeck, S. Toepfer, E. v. Someren, and R. Guthke, "Gene regulatory network inference: Data integration in dynamic models - A review," *BioSystem*, vol. 96, pp. 86-103, 2009.
- [15] R. Albert, "Network inference, analysis, and modeling in systems biology," *The Plant Cell*, vol. 19, pp. 3327-3338, 2007.
- [16] D. Anastassiou, "Computational analysis of the synergy among multiple interacting genes," *Molecular Systems Biology*, vol. 3, pp. 83-83, 2007.
- [17] K. J. Friston, "Causal modelling and brain connectivity in functional magnetic resonance imaging," *PLoS Biology*, vol. 7, p. e1000033, 2009.
- [18] C. J. Needham, J. R. Bradford, A. J. Bulpitt, and D. R. Westhead, "A primer on learning in bayesian networks for computational biology," *PLoS Computational Biology*, vol. 3, p. e129, 2007.

- [19] I. Cantone, L. Marucci, F. Iorio, M. A. Ricci, V. Belcastro, M. Bansal, *et al.*, "A yeast synthetic network for in-vivo assessment of reverse-engineering and modeling approaches," *Cell*, vol. 137, pp. 172-181, 2009.
- [20] D. M. Camacho and J. J. Collins, "Systems biology strikes gold," *Cell*, vol. 137, pp. 24-26, 2009.
- [21] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424-438, 1969.
- [22] S. Mukherjee and T. P. Speed, "Network inference using informative priors," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 14313-14318, 2008.
- [23] N. Wiener, "The theory of prediction," *Modern Mathematics for the Engineer*, pp. 165-190, 1956.
- [24] D. A. Pierce and L. D. Haugh, "Causality in temporal systems: Characterization and a survey," *Journal of Econometrics*, vol. 5, pp. 265-293, 1977.
- [25] J. Geweke, "Inference and causality in economic time series models," in *Handbook of Econometrics*. vol. Volume 2, ed: Elsevier, 1984, pp. 1101-1144.
- [26] H. Lütkepohl, *New introduction to multiple time series analysis*: Springer Berlin Heidelberg, 2005.
- [27] D. Tjøstheim, "Granger-causality in multiple time series," *Journal of Econometrics*, vol. 17, pp. 157-176, 1981.

- [28] S. D. Bay, L. Chrisman, A. Pohorille, and J. Shrager, "Temporal aggregation bias and inference of causal regulatory networks," *Journal of Computational Biology*, vol. 11, pp. 971-985, 2004.
- [29] G. Michailidis and F. d'Alché-Buc, "Autoregressive models for gene regulatory network inference: sparsity, stability and causality issues," *Mathematical Biosciences*, vol. 246, pp. 326-34, 2013.
- [30] G. H. F. Tam, C. Chang, and Y. S. Hung, "Application of Granger causality to gene regulatory network discovery," in *6th International Conference on Systems Biology (ISB) 2012*, pp. 232-239.
- [31] S. Ray, M. Haney, C. Hanson, B. Biswal, and S. J. Hanson, "Modeling causal relationship between brain regions within the drug-cue processing network in chronic cocaine smokers," *Neuropsychopharmacology*, 2015.
- [32] H.-S. Choi, Y. Kim, K.-H. Cho, and T. Park, "Integrative analysis of time course microarray data and DNA sequence data via log-linear models for identifying dynamic transcriptional regulatory networks," *International Journal of Data Mining and Bioinformatics*, vol. 7, pp. 38-57, 2012.
- [33] G. Karlebach and R. Shamir, "Modelling and analysis of gene regulatory networks," *Nature Reviews Molecular Cell Biology*, vol. 9, pp. 770-780, 2008.
- [34] A. Fujita, J. Sato, H. Garay-Malpartida, R. Yamaguchi, S. Miyano, M. Sogayar, *et al.*, "Modeling gene expression regulatory networks with the sparse vector autoregressive model," *BMC Systems Biology*, vol. 1, p. 39, 2007.

- [35] E. W. Lang, A. M. Tomé, I. R. Keck, J. M. Górriz-Sáez, and C. G. Puntonet. (2012, Brain connectivity analysis: A short survey. *Computational Intelligence and Neuroscience 2012*, 21.
- [36] W. Liao, J. Ding, D. Marinazzo, Q. Xu, Z. Wang, C. Yuan, *et al.*, "Small-world directed networks in the human brain: multivariate Granger causality analysis of resting-state fMRI," *NeuroImage*, vol. 54, pp. 2683-2694, 2011.
- [37] M. Ding, Y. Chen, and S. L. Bressler, *Granger causality: Basic theory and application to neuroscience*: Wiley, 2006.
- [38] M. Hu and H. Liang, "A copula approach to assessing Granger causality," *NeuroImage*, vol. 100, pp. 125-134, 2014.
- [39] Z. Zhenyu, J. Yun, T. Tianyu, L. Zuhong, I. Yijun, C. Yonghong, *et al.*, "Detecting effective connectivity in human brain using Granger causality," in *International Conference on BioMedical Engineering and Informatics*, 2008, pp. 394-398.
- [40] T. Changa, R. Guptab, R. Inglesi-Lotzb, B. Simo-Kengneb, D. Smithersb, and A. Tremblingb, "Renewable energy and growth: Evidence from heterogeneous panel of G7 countries using Granger causality," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1405-1412, 2015.
- [41] P. Geiger, K. Zhang, B. Schoelkopf, M. Gong, and D. Janzing, "Causal inference by identification of vector autoregressive processes with hidden components," in *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, 2015, pp. 1917-1925.

- [42] Y. Wolde-Rufael, "Electricity consumption and economic growth in transition countries: A revisit using bootstrap panel Granger causality analysis," *Energy Economics*, vol. 44, pp. 325-330, 2014.
- [43] F. Emirmahmutoglu, M. Bacilar, N. Apergis, B. D. Simo-Kengne, T. Chang, and R. Gupta, "Causal relationship between asset prices and output in the United States: Evidence from the state-level panel Granger causality test," *Regional Studies*, pp. 1-14, 2015.
- [44] N. W. Falasca, S. D'Ascenzo, A. D. Domenico, M. Onofrj, L. Tommasi, B. Laeng, *et al.*, "Hemispheric lateralization in top-down attention during spatial relation processing: a Granger causal model approach," *European Journal of Neuroscience*, vol. 41, pp. 914-924, 2015.
- [45] Y. Wang, H. Chen, Q. Gao, Y. Yang, Q. Gong, and F. Gao, "Evaluation of net causal influences in the circuit responding to premotor control during the movement-readiness state using conditional Granger causality," *Brain Research*, vol. 1595, pp. 110-119, 2015.
- [46] G.-R. Wu, W. Liao, S. Stramaglia, and D. Marinazzo, "Recovering directed networks in neuroimaging datasets using partially conditioned Granger causality," *BMC Neuroscience*, vol. 14, p. P260, 2013.
- [47] W. A. Freiwald, P. Valdes, J. Bosch, R. Biscay, J. C. Jimenez, L. M. Rodriguez, *et al.*, "Testing non-linearity and directedness of interactions between neural groups in the macaque inferotemporal cortex," *Journal of Neuroscience Methods*, vol. 94, pp. 105-119, 1999.

- [48] B. Gourévitch, R. L. Bouquin-Jeannès, and G. Faucon, "Linear and nonlinear causality between signals: methods, examples and neurophysiological applications," *Biological Cybernetics*, vol. 95, pp. 349-369, 2006.
- [49] K. Hlavácková-Schindler and H. Bouzari, "Granger LASSO causal models in higher dimensions-application to gene expression regulatory networks," in *ECML/PKDD 2013 Workshop Scalable Decision Making: Uncertainty, Imperfection, Deliberation (SCALE)*, Prague, Czech Republic, 2013.
- [50] Z. G. Zhang, Y. S. Hung, S. C. Chan, W. C. Xu, and Y. Hu, "Modelling and identification of gene regulatory networks: A Granger casuality approach," in *Ninth International Conference on Machine Learning and Cybernetics*, Qingdao, 2010, pp. 3073-3078.
- [51] G. H. F. Tam, Y. S. Hung, and C. Chang, "Meta-analysis on gene regulatory networks discovered by pairwise Granger causality," in *The 7th International Conference on Systems Biology (ISB)* Huangshan, China, 2013, pp. 123-128.
- [52] S. Yao, S. Yoo, and D. Yu, "Prior knowledge driven Granger causality analysis on gene regulatory network discovery," *BMC Bioinformatics*, vol. 16, p. 273, 2015.
- [53] J. E. Larvie, M. S. Gorji, and A. Homaifar, "Inferring stable gene regulatory networks from steady-state data," in *41st Annual Northeast Biomedical Engineering Conference (NEBEC)* 2015, pp. 1-2.
- [54] A. Fujita, J. R. Sato, K. Kojima, L. R. Gomes, M. Nagasaki, M. C. Sogayar, *et al.*, "Identification of Granger causality between gene sets," *Journal of Bioinformatics and Computational Biology*, vol. 8, pp. 679-701, 2010.

- [55] S. Guo, A. K. Seth, K. M. Kendrick, C. Zhou, and J. Feng, "Partial Granger causality—Eliminating exogenous inputs and latent variables," *Journal of Neuroscience Methods*, vol. 172, pp. 79-93, 2008.
- [56] A. Fujita, P. Severino, J. R. Sato, and S. Miyano, "Granger causality in systems biology: Modeling gene networks in time series microarray data using vector autoregressive models," in *Advances in Bioinformatics and Computational Biology*. vol. 6268, C. E. Ferreira, S. Miyano, and P. F. Stadler, Eds., ed: Springer Berlin Heidelberg, 2010, pp. 13-24.
- [57] S. Guo, J. Wu, M. Ding, and J. Feng, "Uncovering interactions in the frequency domain," *PLoS Computational Biology*, vol. 4, p. e1000087, 2008.
- [58] A. Fujita, A. Patriota, J. Sato, and S. Miyano, "The impact of measurement errors in the identification of regulatory networks," *BMC Bioinformatics*, vol. 10, p. 412, 2009.
- [59] M. B. Priestley, *Non-linear and non-stationary time series analysis*. London: Academic Press, 1988.
- [60] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, p. 461, 2000.
- [61] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, "A robust method for detecting interdependences: application to intracranially recorded EEG," *Physica D: Nonlinear Phenomena*, vol. 134, pp. 419-430, 1999.

- [62] R. Q. Quiroga, J. Arnhold, and P. Grassberger, "Learning driver-response relationships from synchronization patterns," *Physical Review E*, vol. 61, p. 5142, 2000.
- [63] R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, "Performance of different synchronization measures in real data: A case study on electroencephalographic signals," *Physical Review E*, vol. 65, p. 041903, 2002.
- [64] J. Bhattacharya, E. Pereda, and H. Petsche, "Effective detection of coupling in short and noisy bivariate data," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 33, pp. 85-95, 2003.
- [65] M. Breakspear and J. R. Terry, "Topographic Organization of Nonlinear Interdependence in Multichannel Human EEG," *NeuroImage*, vol. 16, pp. 822-835, 2002.
- [66] M. Wiesenfeldt, U. Parlitz, and W. Lauterborn, "Mixed state analysis of multivariate time series," *International Journal of Bifurcation and Chaos*, vol. 11, pp. 2217-2226, 2001.
- [67] Q. Luo, T. Ge, and J. Feng, "Granger causality with signal-dependent noise," *NeuroImage*, vol. 57, pp. 1422-1429, 2011.
- [68] A. K. Seth, "A MATLAB toolbox for Granger causal connectivity analysis," *Journal of Neuroscience Methods*, vol. 186, pp. 262-273, 2010.

- [69] M. Y. Siyal, M. S. Furqan, and S. M. G. Monir, "Granger causality: Comparative analysis of implementations for Gene Regulatory Networks," in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, 2014, pp. 793-798.
- [70] Z. Reitermanova, "Data splitting," *WDS'10 Proceedings of Contributed Papers*, vol. Part 1, pp. 31-36, 2010.
- [71] A. Zheng, "Offline evaluation mechanisms: Hold-out validation, cross-validation, and bootstrapping," in *Evaluating machine learning models: A beginner's guide to key concepts and pitfalls*, S. Cutt, Ed., ed: O'Reilly Media, 2015.
- [72] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning : data mining, inference, and prediction*: New York : Springer, c2009.;2nd ed., 2009.
- [73] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716-723, 1974.
- [74] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461-464, 1978.
- [75] B. Efron, "Bootstrap methods: Another look at the jackknife," *The Annals of Statistics*, vol. 7, pp. 1-16, 1979.
- [76] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, pp. 79-86, 1951.

- [77] D. R. Anderson, K. P. Burnham, and G. C. White, "Kullback-Leibler information in resolving natural resource conflicts when definitive data exist," *Wildlife Society Bulletin*, vol. 29, pp. 1260-1270, 2001.
- [78] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: A practical information-theoretic approach*, 2 ed.: Springer-Verlag New York, 2002.
- [79] M. J. Mazerolle, "Improving data analysis in herpetology: using Akaike's Information Criterion (AIC) to assess the strength of biological hypotheses," *Amphibia-Reptilia*, vol. 27, pp. 169-180, 2006.
- [80] C. Giraud, *Introduction to high-dimensional statistics*: Chapman and Hall/CRC, 2014.
- [81] L. A. Jaekel. (1972) The infinitesimal Jackknife. *Memorandum for File*.
- [82] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society, Series B*, vol. 67, pp. 301-320, 2005.
- [83] A. S. A. Goldberger, *Econometric Theory*: John Wiley & Sons, 1964.
- [84] T. Shimamura, S. Imoto, R. Yamaguchi, A. Fujita, M. Nagasaki, and S. Miyano, "Recursive regularization for inferring gene networks from time-course gene expression profiles," *BMC Systems Biology*, vol. 3, p. 41, 2009.
- [85] Z. Li and M. J. Sillanpaa, "Overview of LASSO - related penalized regression methods for quantitative trait mapping and genomic selection," *Theoretical and Applied Genetics*, vol. 125, pp. 419-435, 2012.

- [86] X. Wang and C. Leng, "High dimensional ordinary least squares projection for screening variables," *Statistical Methodology, Series B* vol. 78, pp. 589-611, 2016.
- [87] C. Ugrinowitsch, G. W. Fellingham, and M. D. Ricard, "Limitations of ordinary least squares models in analyzing repeated measures data," *Medicine and Science in Sports and Exercise*, vol. 36, pp. 2144-2148, 2004.
- [88] K. Hlaváčková-Schindler and S. J. Pereverzyev, "Lasso Granger causal models: Some strategies and their efficiency for gene expression regulatory networks," in *Decision making: Uncertainty, imperfection, deliberation and scalability*. vol. 538, T. V. Guy, M. Kárný, and D. H. Wolpert, Eds., ed: Springer International Publishing, 2015, pp. 91-117.
- [89] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 58, pp. 267-288, 1996.
- [90] R. Tibshirani, "The LASSO method for variable selection in the COX model," *Statistics in Medicine*, vol. 16, pp. 385-395, 1997.
- [91] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55-67, 1970.
- [92] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, pp. 1418-1429, 2006.

- [93] J. N. Morgan and J. A. Sonquist, "Problems in the Analysis of Survey Data, and a Proposal," *Journal of the American Statistical Association*, vol. 58, pp. 415-434, 1963.
- [94] J. A. Sonquist, *Multivariate model building; the validation of a search strategy* Ann Arbor : Survey Research Center, University of Michigan, 1970.
- [95] R. Messenger and L. Mandell, "A modal search technique for predictive nominal scale multivariate analysis," *Journal of the American Statistical Association*, vol. 67, pp. 768-772, 1972.
- [96] M. W. Gillo, "MAID: A Honeywell 600 program for an automated survey analysis," *Behavioral Science*, vol. 17, pp. 251-252, 1972.
- [97] M. W. Gillo and M. W. Shelly, "Predictive modeling of multivariable and multivariate data," *Journal of the American Statistical Association*, vol. 69, pp. 646-653, 1974.
- [98] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression tree*: Champman and Hall/CRC, 1984.
- [99] J. H. Friedman, *A tree-structured approach to nonparametric multiple regression*. Berlin: Springer Verlag, 1979.
- [100] J. R. Quinlan, *C4.5: Programs for Machine Learning*: Morgan Kaufmann Publishers, Inc., 1993.
- [101] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81-106, 1986.

- [102] J. R. Quinlan, "Discovering rules by induction from large collections of examples," in *Expert systems in the micro electronic age*, D. Michie, Ed., ed: Edinburgh University Press, 1979.
- [103] M. H. Quenouille, "Approximate tests of correlation in time-series," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 11, pp. 68-84, 1949.
- [104] P. J. Bickel and D. A. Freedman, "Some asymptotic theory for the bootstrap," *The Annals of Statistics*, vol. 9, pp. 1196-1217, 1981.
- [105] K. Singh, "On the asymptotic accuracy of Efron's bootstrap," *The Annals of Statistics*, vol. 9, pp. 1187-1195, 1981.
- [106] T. J. DiCiccio and B. Efron, "Bootstrap confidence intervals," *Statistical Science*, vol. 11, pp. 189-228, 1996.
- [107] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, pp. 123-140, 1996.
- [108] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.
- [109] A. Liaw and M. Wiener, "Classification and regression by random forest," *R News*, vol. 2, pp. 18-22, 2002.
- [110] J. Liu, S. Ji, and J. Ye, "SLEP: Sparse learning with efficient projections," 4.1 ed: Arizona State University, 2009.
- [111] L. Barnett and A. K. Seth, "The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference," *Journal of Neuroscience Methods*, vol. 223, pp. 50-68, 2014.

- [112] B. Schelter, M. W. M. Eichler, M. Peifer, B. Hellwig, B. Guschlbauer, C. H. Lücking, *et al.*, "Testing for directed influences among neural signals using partial directed coherence," *Journal of Neuroscience Methods*, vol. 152, pp. 210-219, 2006.
- [113] N. D. Mukhopadhyay and S. Chatterjee, "Causality and pathway search in microarray time series experiment," *Bioinformatics*, vol. 23, pp. 442-449, 2007.
- [114] G. H. F. Tam, C. Chunqi, and Y. S. Hung, "Gene regulatory network discovery using pairwise Granger causality," *IET Systems Biology*, vol. 7, pp. 195-204, 2013.
- [115] L. A. Baccala and K. Sameshima, "Partial directed coherence: A new concept in neural structure determination," *Biological Cybernetics*, vol. 84, pp. 463-474, 2001.
- [116] E. AccessScience. (2014, Importance of HeLa cells. Available: <http://www.accessscience.com/content/importance-of-hela-cells/BR0826141>
- [117] M. L. Whitfield, G. Sherlock, A. J. Saldanha, J. I. Murray, C. A. Ball, K. E. Alexander, *et al.*, "Identification of genes periodically expressed in the human cell cycle and their expression in tumors," *Molecular Biology of the Cell*, vol. 13, pp. 1977-2000, 2002.
- [118] M. L. Whitfield, G. Sherlock, A. Saldanha, J. I. Murray, C. A. Ball, K. E. Alexander, *et al.* (2002). *Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors*. Available: <http://genome-www.stanford.edu/Human-CellCycle/HeLa/>

- [119] A. C. Lozano, N. Abe, Y. Liu, and S. Rosset, "Grouped graphical Granger modeling for gene expression regulatory networks discovery," *Bioinformatics*, vol. 25, pp. i110-8, 2009.
- [120] P. J. Green and B. W. Silverman, *Nonparametric regression and generalized linear models: A roughness penalty approach*. New York: Chapman and Hall, 1994.
- [121] X. Li, S. Rao, W. Jiang, C. Li, Y. Xiao, Z. Guo, *et al.*, "Discovery of time-delayed gene regulatory networks based on temporal gene expression profiling," *BMC Bioinformatics*, vol. 7, pp. 1-20, 2006.
- [122] T. A. Keller, M. A. Just, and V. A. Stenger, "Reading span and the time-course of cortical activation in sentence-picture verification," presented at the Annual Convention of the Psychonomic Society, Orlando, FL, 2001.
- [123] T. Mitchell and W. Wang. (2001). *StarPlus fMRI data*. Available: <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>
- [124] T. Mitchell, R. Hutchinson, R. S. Niculescu, F. Pereira, and X. Wang, "Learning to decode cognitive states from brain images," *Machine Learning*, vol. 57, pp. 145-175, 2004.
- [125] X. Wang and T. Mitchell, "Detecting cognitive states using machine learning," *Iterim working paper*, 2002.

[

- [126] R. A. Hutchinson, R. S. Niculescu, T. A. Keller, I. Rustandi, and T. M. Mitchell, "Modeling fMRI data generated by overlapping cognitive processes with unknown onsets using hidden process models," *NeuroImage*, vol. 46, pp. 87-104, 2009.
- [127] W. F. Eddy, M. F. C. Genovese, N. Lazar, AudrisMockus, and J. Welling, "The challenge of fnctional magnetic resonance imaging," *Journal of Computational and Graphical Statistics*, vol. 8, pp. 545-558, 1999.
- [128] Y. Yang, "Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation," *Biometrika*, vol. 92, pp. 937-950, 2005.
- [129] A. Nemirovski. (1994/95). *Efficient methods in convex programming*. Available: www2.isye.gatech.edu/~nemirovs/Lect_EMCO.pdf
- [130] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, pp. 183-202, 2009.
- [131] Y. Nesterov, "Gradient methods for minimizing composite functions," *Mathematical Programming*, vol. 140, pp. 125-161, 2013.
- [132] J. Liu, S. ji, and J. Ye, "SLEP: Spare Learning with Efficient Projection Manual v 4.1," A. S. University, Ed., ed, 2011.
- [133] J. C. Culham and N. G. Kanwisher, "Neuroimaging of cognitive functions in human parietal cortex," *Current opinion in Neurobiology*, vol. 11, pp. 157-163, 2001.

- [134] J. M. Rumsey, B. Horwitz, B. C. Donohue, K. Nace, J. M. Maisog, and P. Andreason, "Phonological and orthographic components of word recognition A PET-rCBF study," *Brain*, vol. 120, pp. 739-759, 1997.
- [135] J. R. Binder, K. A. McKiernan, M. E. Parsons, C. F. Westbury, E. T. Possing, J. N. Kaufman, *et al.*, "Neural correlates of lexical access during visual word recognition," *Journal of Cognitive Neuroscience*, vol. 15, pp. 372-393, 2003.
- [136] J. D. Edwards, P. M. Pexman, B. G. Goodyear, and C. G. Chambers, "An fMRI investigation of strategies for word recognition," *Cognitive Brain Research*, vol. 24, pp. 648-662, 2005.
- [137] M. W. L. Chee, K. M. O'Craven, R. Bergida, B. R. Rosen, and R. L. Savoy, "Auditory and visual word processing studied with fMRI," *Human Brain Mapping*, vol. 7, pp. 15-28, 1999.
- [138] J. R. Booth, D. D. Burman, J. R. Meyer, D. R. Gitelman, T. B. Parrish, and M. M. Mesulam, "Functional anatomy of intra- and cross-modal lexical tasks," *NeuroImage*, vol. 16, pp. 7-22, 2002.
- [139] S. Dehaene, L. Naccache, L. Cohen, D. L. Bihan, J.-F. Mangin, J.-B. Poline, *et al.*, "Cerebral mechanisms of word masking and unconscious repetition priming," *Nature Neuroscience*, vol. 4, pp. 752-758, 2001.
- [140] M. Vigneau, V. Beaucoisin, P. Y. Hervé, H. Duffau, F. Crivello, O. Houdé, *et al.*, "Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing," *NeuroImage*, vol. 30, pp. 1414-1432, 2006.

- [141] M. Grande, E. Meffert, E. Schoenberger, S. Jung, T. Frauenrath, W. Huber, *et al.*, "From a concept to a word in a syntactically complete sentence: An fMRI study on spontaneous language production in an overt picture description task," *NeuroImage*, vol. 61, pp. 702-714, 2012.
- [142] H. R. Heekeren, S. Marrett, D. A. Ruff, P. A. Bandettini, and L. G. Ungerleider, "Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality," *Proceedings of the National Academy of Sciences of United States of America*, vol. 103, pp. 10023-10028, 2006.
- [143] R. Elliott, "Executive functions and their disorders: Imaging in clinical neuroscience," *British Medical Bulletin*, vol. 65, pp. 49-59, 2003.
- [144] S. Monsell, "Task switching," *Trends in Cognitive Sciences*, vol. 7, pp. 134-140, 2003.
- [145] R. C. K. Chan, D. Shum, T. Touloupoulou, and E. Y. H. Chen, "Assessment of executive functions: Review of instruments and identification of critical issues," *Archives of Clinical Neuropsychology*, vol. 23, pp. 201-2016, 2008.
- [146] A. Chatr-aryamontri, B.-J. Breitkreutz, R. Oughtred, L. Boucher, S. Heinicke, D. Chen, *et al.*, "The BioGRID interaction database: 2015 update," in *Nucleic Acids Research*, ed, 2014.
- [147] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel Granger causality and the analysis of dynamical networks," *Physical Review E*, vol. 77, 2008.

- [148] P. O. Amblard, R. Vincent, O. J. J. Michel, and C. Richard, "Kernelizing Geweke's measures of granger causality," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* 2012, pp. 1-6.
- [149] R. B. Nelsen, *An introduction to copulas*, 2nd ed. New York: Springer-Verlag, 2006.
- [150] H. Liu, J. Lafferty, and L. Wasserman, "The nonparanormal: Semiparametric estimation of high dimensional undirected graphs," *Journal of Machine Learning Research*, vol. 10, pp. 2295-2328, 2009.
- [151] M. T. Bahadori and Y. Liu, "An examination of practical Granger causality inference," in *2013 SIAM International Conference on Data Mining*, Austin, Texas, USA, 2013, pp. 467-475.
- [152] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, pp. 1348-1360, 2001.
- [153] J. H. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, pp. 1-22, 2010.
- [154] J. O. Ogutu, T. Schulz-Streeck, and H.-P. Piepho, "Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions," in *15th European workshop on QTL mapping and marker assisted selection (QTLMAS)*, 2012, p. S10.

- [155] H.-I. Suk, C.-Y. Wee, and D. Shen, "Discriminative group sparse representation for mild cognitive impairment classification," in *Machine Learning in Medical Imaging*. vol. 8184, G. Wu, D. Zhang, D. Shen, P. Yan, K. Suzuki, and F. Wang, Eds., ed: Springer International Publishing, 2013, pp. 131-138.
- [156] J. Liu, Y. Xu, J. Cheng, Z. Zhang, D. W. K. Wong, F. Yin, *et al.*, "Multiple modality fusion for glaucoma diagnosis," in *The International Conference on Health Informatics*. vol. 42, Y.-T. Zhang, Ed., ed: Springer International Publishing, 2014, pp. 5-8.
- [157] Y. Liu, M. T. Bahadori, and H. Li. (2012). *Copula-Granger*. Available: www.scf.usc.edu/~mohammab/codes/codes.html.
- [158] J. Qian, T. Hastie, J. Friedman, R. Tibshirani, and N. Simon. (2013). *Glmnet for MATLAB*. Available: http://www.stanford.edu/~hastie/glmnet_matlab/
- [159] D. Cheng, M. T. Bahadori, and Y. Liu, "FBLG: a simple and effective approach for temporal dependence discovery from time series data," presented at the Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, New York, USA, 2014.
- [160] R. Wolthuis, L. Clay-Farrace, W. v. Zon, M. Yekezare, L. Koop, J. Ogink, *et al.*, "Cdc20 and Cks direct the spindle checkpoint-independent destruction of Cyclin A," *Molecular Cell*, vol. 30, pp. 290-302, 2008.

- [161] M. Kong, E. A. Barnes, V. Ollendorff, and D. J. Donoghue, "Cyclin F regulates the nuclear localization of cyclin B1 through a cyclin–cyclin interaction," *The EMBO Journal*, vol. 19, pp. 1378-1388, 2000.
- [162] M. Radulovic, E. Crane, M. Crawford, J. Godovac-Zimmermann, and V. P. C. C. Yu, "CKS proteins protect mitochondrial genome integrity by interacting with mitochondrial single-stranded DNA-binding protein," *Molecular and Cellular Proteomics*, vol. 9, pp. 145-152, 2010.
- [163] S. Haufe, V. V. Nikulin, and G. Nolte, "Alleviating the influence of weak data asymmetries on Granger-causal analyses," in *Latent Variable Analysis and Signal Separation: 10th International Conference, LVA/ICA 2012, Tel Aviv, Israel, March 12-15, 2012. Proceedings*, F. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 25-33.
- [164] M. Hu, W. Li, and H. Liang, "A Copula-based Granger causality measure for the analysis of neural spike train data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. PP, pp. 1-1, 2015.
- [165] M. Eichler, "Graphical modelling of multivariate time series," *Probability Theory and Related Fields*, vol. 153, pp. 233-268, 2011.
- [166] M. T. Bahadori. (2014). *Lasso-Granger*. Available: <http://www-scf.usc.edu/~mohammab/codes/codes.html>
- [167] M. S. Furqan and M. Y. Siyal, "Biological network detection using forward backward pair-wise Granger causality," *International Journal of Simulation-Systems, Science and Technology*, vol. 17, 2016.

- [168] M. S. Furqan and M. Y. Siyal, "Inference of biological networks using bi-directional random forest Granger causality," *SpringerPlus*, vol. 5, p. 514, 2016.