3-21-2013

Network Intrusion Dataset Assessment

David J. Weller-Fahy

NETWORK INTRUSION DATASET ASSESSMENT

THESIS

David J. Weller-Fahy, Senior Master Sergeant, USAF

AFIT-ENG-13-M-49

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-13-M-49

NETWORK INTRUSION DATASET ASSESSMENT

THESIS

Presented to the Faculty

Department of Electrical and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cyber Operations

David J. Weller-Fahy, B.S.C.S.

Senior Master Sergeant, USAF

March 2013

AFIT-ENG-13-M-49

NETWORK INTRUSION DATASET ASSESSMENT

David J. Weller-Fahy, B.S.C.S.
Senior Master Sergeant, USAF

Approved:

Brett J. Borghetti, Ph.D. (Chairman)	7 MAR 2013 Date
Angela A. Sodemann, Ph.D. (Member)	3/6/2013 Date
Rusty O. Baldwin, Ph.D. (Member)	7 Mar 13 Date

AFIT-ENG-13-M-49

Abstract

Research into classification using Anomaly Detection (AD) within the field of Network Intrusion Detection (NID), or Network Intrusion Anomaly Detection (NIAD), is common, but operational use of the classifiers discovered by research is not. One reason for the lack of operational use is most published testing of AD methods uses artificial datasets: making it difficult to determine how well published results apply to other datasets and the networks they represent. This research develops a method to predict the accuracy of an AD-based classifier when applied to a new dataset, based on the difference between an already classified dataset and the new dataset. The resulting method does not accurately predict classifier accuracy, but does allow some information to be gained regarding the possible range of accuracy. Further refinement of this method could allow rapid operational application of new techniques within the NIAD field, and quick selection of the classifier(s) that will be most accurate for the network.

For my wife — you provide the vector.

Acknowledgments

I would like to express my gratitude to my advisers, Dr. Angela Sodemann and Lt Col Brett Borghetti, for their insight, criticism, and patience with my brainstorm process throughout the course of this research. Their help, experience, and knowledge was invaluable.

I would like to thank my thesis committee member Dr. Rusty Baldwin for his thoughtful critiques and analysis throughout my research. I would also like to thank Dr. Gilbert Peterson for insight into classification schemes, and Lt Col Jeffrey Clark for brainstorming and providing the blade to solve some Gordian Knots.

I would like to thank Dr. D'Amico for setting me on a path to statistical relevance and humoring someone who didn't know which questions to ask.

Finally, I would like to thank my children who will no longer have to ask why I work every weekend.

David J. Weller-Fahy

Table of Contents

Page

List of Figures

List of Tables

List of Acronyms

Acronym	Definition
AD	Anomaly Detection 1
AN	Actual Negatives 38
AP	Actual Positives 38
BCT	Binary Classification Tree 53
CAPS	Classifier Accuracy Prediction System 58
CUT	Component Under Test 29
DARPA	Defense Advanced Research Projects Agency 9
FPR	False Positive Rate 52
FP	False Positives 38
ID	Intrusion Detection 6
IDEVAL	Intrusion Detection Evaluation 9
IDS	Intrusion Detection System 11
KDD99	1999 Knowledge Discovery and Data Mining Tools Competition 30
LDA	Linear Discriminant Analysis 36
NB	Naïve Bayes 53
NIAD	Network Intrusion Anomaly Detection 1
NIDS	Network Intrusion Detection System 8
NID	Network Intrusion Detection 1
NI	Network Intrusion 59
QDA	Quadratic Discriminant Analysis 52
QQ	Quantile-Quantile 46
SUT	System Under Test 28
SVM	Support Vector Machine 42

Acronym	Definition
TN	True Negatives 26
TPR	True Positive Rate 52
TP	True Positives 38

NETWORK INTRUSION DATASET ASSESSMENT

I. Introduction

Network Intrusion (NI) refers to a myriad of techniques and technologies that can be used to penetrate and exploit computer networks. NI datasets are collections of network traffic that evaluate new Network Intrusion Detection (NID) classifiers. There are two types of NID classifiers: Anomaly Detection (AD) and misuse.

Recently, there has been significant focus on attacks vectored through gaps in network security, leading to renewed interest in the effectiveness of intrusion prevention measures. However, to prevent intrusion there needs to be an effective method of Intrusion Detection (ID). As misuse detection is unable to detect novel attacks, the focus has rested largely on applications of AD to the NID field, or Network Intrusion Anomaly Detection (NIAD).

1.1 Problem Statement

The objective of this research is to develop a system by which the differences measured between networks can be used to predict the corresponding change in classifier performance. In the NIAD field networks are examined via traffic captured from the network, which is then converted into a dataset by generating features from that traffic. The use of datasets to attempt to examine networks is widespread and accepted, therefore this research will use the terms network and dataset interchangeably.

The lack of a standard method to compare datasets (thus networks) means there is little connection between the research and real-world systems: most current research is inapplicable to operational systems without expensive testing. While methods are available for measuring the theoretical upper-bound on the capability of a Network Intrusion Detection

System (NIDS) [22], there is little information published in the field on the use of datasets to evaluate NIDS. NIDS developers could use a method of predicting the change in classifier performance between datasets as a metric when developing new classifiers.

Specifically, the goals of this research are to:

- develop a structure suitable for conducting experiments on the correlations of the difference between datasets and the corresponding change in classifier performance,
- identify difference measures which correlate to changes in classifier performance,
- build and evaluate a system to predict the difference in classifier accuracy between two datasets.

All goals were achieved by this research effort. Analysis identified multiple difference measures that correlated to various classifier performance measures, and the model developed during research is able to predict classifier performance change based on differences between datasets.

1.2 Contributions

This research provides two significant contributions to the field of NIAD. First, this research demonstrates consistent correlation of the differences between datasets and the changes in corresponding classifier performance measures. While the correlations found were not high, they demonstrate a relationship between the difference measures used in this research and the performance of a wide range of classifiers.

Second, this research provides a new way of approaching the, "No free lunch theorem," proposed by Wolpert and Macready [61] as it may apply to finding an optimal classifier for a given dataset. Approaching the problem of selecting a classifier for a particular dataset as an optimization problem, and trying to predict the outcome of the classifier using the differences between datasets, provides a new framework within which to approach the problem of classifier selection.

1.3 Overview

Chapter 2 examines the topics surrounding NIAD and summarizes the phases of ID. It then provides an overview of the available NI datasets along with a discussion of the different types of datasets available. Once the available datasets are covered, methods of dataset characterization are discussed leading into the last section in Chapter 2. The last section reviews current research in the area of distance and similarity measures, along with their possible uses in the NIAD field.

Chapter 3 describes the Classifier Accuracy Prediction System (CAPS) that is developed to perform experiments on dataset differences and changes in classifier performance. The parameters and boundaries of the CAPS are examined, as well as the factors and levels of the experiment. Finally, discussion proceeds on how the experimental results are analyzed and evaluated.

Chapter 4 reports the results of the three sets of experiments designed to accomplish the goals set forth in Section 1.1. Analysis of the correlation within and across datasets is performed, with visual checks occurring after the statistical analysis is complete. The predictive model is developed and evaluated, and the results of that evaluation are reported.

Chapter 5 summarizes the experimental results, and provides an evaluation of the quality of the predictive model. Based on the provided evidence, the conclusion is that there is correlation of the difference between datasets and the corresponding change in classifier performance. In addition to correlation, linear models are developed which are able to predict accuracy across all classifiers, but not True Positive Rate (TPR) or False Positive Rate (FPR).

II. Background

2.1 Related Work

The successful characterization of a Network Intrusion (NI) dataset, and of the difference between two NI datasets, rests largely on how the differences and similarities between the two datasets are quantified. In particular, the method of measuring the distance or similarity between two datasets is critical to successful determination of whether two datasets are similar in structure and content, and whether results of an evaluation on one may apply to the other.

Within the Network Intrusion Anomaly Detection (NIAD) field the closest work to this research is, "On the distance norms for detecting anomalies in multidimensional datasets," by Chmielewski and Wierzchoń [9] which examines the problems inherent in using a form of power (p, r)-distance. Power (p, r)-distance measures the distance between two vectors x and y of length n.

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{r}} \tag{2.1}$$

The particular distance used is determined by the values of p and r, where the values assigned to each will sometimes result in distances that may be familiar to readers. In example, where $p = r = 2$ the power (p, r)-distance is more commonly known as the Euclidean metric, and $p = r \geq 1$ the power (p, r)-distance is known more generally as the l_p-metric. Chmielewski and Wierzchoń use the l_p-metric (Equation (2.1), $p = r \geq 1$), fractional l_p-distance (Equation (2.1), $0 < p = r < 1$), and cosine similarity (Equation (2.2)) to measure the distance between different samples of high-dimensional data. Cosine similarity is shown in Equation (2.2), where ϕ is the angle between vectors x and y:

$$\cos \phi = \frac{\langle x, y \rangle}{\sqrt{x^2} \cdot \sqrt{y^2}} \tag{2.2}$$

Through experimentation using differing values of p on the l_p-metric, and using the resulting distance in an application of negative selection to a NI dataset, they conclude that values of p on the interval $[0.5, 1.0]$ should provide an improvement in detection rate compared to other values. While they do not address the issues of correlation between difference and classifier performance measures, the examination of different forms of distance is useful.

Outside the NIAD field the closest work to this research is, "Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions," by Cha [6] which provides a syntactic and semantic categorization of distance and similarity measures as applied to probability distribution functions, as well as an analysis of the correlation between different measures using clustering and presented in hierarchical clusters. While not a review of how the measures are used, Cha's work is a useful reference for distance measures with an interesting partitioning based on how well the distance measures correlate to each other. A similar work with different intent is the, "Encyclopedia of Distances," by Deza and Deza [15] which provides a comprehensive enumeration of the main distance measures used within a variety of different fields. The cross-disciplinary manner in which the list of distance measures is treated is especially useful when trying to identify measures used in published works, as synonyms and similar formulations are referenced throughout.

In areas not directly related to measuring difference and similarity in multi-dimensional datasets, but which may be useful in examining multi-dimensional datasets, there are other significant works. Within, "Ontologies and Similarity," Staab [52] examines the relationships between ontologies and similarity measures, especially in light of the use of measures within logical reasoning systems. Cunningham [11] develops, "A Taxonomy of Similarity Mechanisms for Case-Based Reasoning," which provides a useful structure for reasoning about which similarity measures to use when first examining a problem.

2.2 Intrusion Detection Phases

Before discussing the NIAD field, it is useful to ensure common understanding of the terms being used. As discussion herein is about the phases of Intrusion Detection (ID) and datasets collected, defining the phases as used within this research is useful. Definitions of the terms used to refer to those phases within this research follow.

- *Preprocessing*: Manipulation of the dataset required to allow the authors' tools to operate on the dataset that is presumed to have no effect on the outcome of the experiment. For example, conversion from the comma-separated-value format to a database table within a relational database may be required, leaving the values of the features within the dataset unchanged.

- *Feature generation*: Creation of new features based on original or derived datasets. For example, conversion of a feature with seven possible categorical values to seven binary features.

- *Selection*: Selection of a subset of all available features or observations for use in classification.

- *Classification*: Categorization of samples as a particular class.

2.3 Network Intrusion Datasets

Careful dataset creation can allow NIAD classifiers to be tested against NI datasets which are representative of specific networks, thus enabling quantifiable comparison with respect to a particular network. However, the difficulties involved in collecting and sharing network traffic [31] have prevented, thus far, the creation of any recent and widely-accepted field-wide standard datasets [16] or standard methods by which NIAD classifiers may be evaluated [32]. The lack of standard datasets and methods engenders questions of how Network Intrusion Detection (NID) are compared, what datasets are used in the comparisons, and whether the comparisons are valid.

A challenge to research into the state of NI datasets is the definition of the term *dataset*. The meaning of the term *dataset* differs among fields of study. In the context of NI, *dataset* can mean a collection of data from any number of different sources, from host system log files to raw packet captures. To refine the scope of this review, in this paper the term *dataset* will refer to a collection of captured packets from live network traffic and any resultant meta-data (such as flow information), *or* a collection of rules to generate packets representing network traffic. Application level data (such as logs) and related analysis tools are not considered as part of this review.

There is a lack of comprehensive recent surveys in the area of NI datasets. The most recent work, "Public domain datasets for optimizing NI and machine learning approaches," by Deraman et al. [14] focused on the availability of public domain datasets and repositories, rather than the qualities needed in future benchmark datasets. The authors identify the need for researchers to share datasets, and discuss how quality datasets would prove useful in NI research. The work closest to a review , the paper "Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods," by Tavallaee et al. [56] is focused on the state of the art in Anomaly Detection (AD) methods, and in identifying pitfalls encountered in published works between 2000 and 2008. They examine 276 studies and encounter a wide variety of problems within the studies. The problems can be summarized as relating to dataset choice and use, poor experimental practices, ineffective presentation of results, and lack of consideration for efficiency in the proposed methods.

Another work close to this review, "Uses and Challenges for Network Datasets," by Heidemann and Papdopoulos [24] is more general in scope. The authors address current research and open questions about network traffic and topology, and which classes of data are collected and used. They examine lessons learned from their research in the areas of privacy and anonymization, and how validation of NID approaches require weak data anonymization. Repeated collection of data is suggested as a way to maintain available

datasets relevant to modern network traffic patterns. They also suggest that multiple datasets of the same type be collected and released to provide the ability to cross-check data. Finally, they describe future work in improving anonymization of captured data, understanding attacks on privacy, capturing and managing annotation and meta-data, focusing on datasets collected from other access types, and developing best practices to deal with the social and legal tension between research and privacy. Heidemann and Papdopoulos were thorough in their coverage of network traffic datasets, and their conclusions apply equally to NI datasets (a subset of network traffic datasets).

Other related studies have been done in validating measurement-based networking research [32], developing an information-theoretic model to describe and evaluate NID systems [22], devising a new metric with which to measure intrusion detection capability [21], surveying the available anomaly-based NID techniques and systems [18], and lessons from technical to legal learned while documenting dataset collection methodology [41].

There are few datasets available that are used in NIAD classifier performance validation. The lack of available datasets is surprising given that the amount of recent work on NID has inspired multiple recent surveys of the subject area: Sommer and Paxson [48] examine the problems encountered when using machine learning to perform NIAD, and set guidelines intended to strengthen AD research. Davis and Clark [12] review the techniques being used in AD preprocessing, and concludes that deep packet inspection and the features derived thereby are required to detect current attacks. Deb et al. [13] cover the NIAD state of the art in wireless ad-hoc and mesh networks, and determine that more work is needed in development of scalable cross-layer Network Intrusion Detection System (NIDS) in the wireless domain. Jyothsna et al. [29] survey the full NIAD field giving an overview of the different methods used to detect anomalous patterns in network traffic. Finally, Vigna [59] examines the history of NID research, and identifies the important role context will play in future NID research. The volume of recent work indicates interest in the field, but research

into new approaches and validation of new NIAD classifiers tends to use one particular dataset: 1999 Knowledge Discovery and Data Mining Tools Competition (KDD99) dataset [53].

2.3.1 1999 Knowledge Discovery and Data Mining Tools Competition Dataset.

The KDD99 dataset [35] is derived from raw packet data generated during a 1998 Intrusion Detection Evaluation (IDEVAL), and contains seven million connection records each with 41 distinct features. The IDEVAL was performed for the Defense Advanced Research Projects Agency (DARPA) by the Massachusetts Institute of Technology's Lincoln Laboratory. Since then, it has become the de facto standard used in research on new NIAD approaches. Multiple studies have analyzed the utility of the KDD99 dataset for NIAD evaluation, but conclusions as to its use as a benchmark dataset vary: Cho et al. [10] recommend not using the KDD99 dataset at all, while Engen et al. [16] suggest that more care be taken in interpretation of results, but recommend continued use. As discussed by Engen et al. [16], researchers continue to use the KDD99 dataset despite its problems for two reasons: First, there is currently no better alternative freely available. Second, there is a large body of work already based on the KDD99 dataset, thus new research may be compared to the existing body of knowledge.

The source of the KDD99 dataset is the 1998 DARPA IDEVAL [36]. The IDEVAL datasets are the largest completely labeled network attack datasets publicly available as full packet captures. The IDEVAL datasets contain a total of 10 weeks of training data, four weeks of test data, and sample file system dumps. In total, the IDEVAL datasets are composed of three distinct datasets — the 1998 evaluation data, the 1999 evaluation data, and scenario specific runs done in 2000. Despite the comprehensiveness of the IDEVAL datasets, their validity has been questioned by some researchers. McHugh [38] found significant problems with the IDEVALs structure, documentation, and resultant data; Mahoney and Chan [37] found artifacts within the data, such as the time-to-live values for all attack packets

differing from the values for all normal packets; and Brugger [4] found that the IDEVAL datasets' background traffic did not emulate normal background traffic. While these flaws impact the use of the IDEVAL datasets in NIAD analysis and validation, it remains the only comprehensive fully labeled benchmark dataset in the field [16] and, thus, widely used.

2.3.2 NSL-KDD Dataset.

To correct the problems identified with the KDD99 dataset, Tavallaee et al. [55] created the NSL-KDD dataset by removing redundant records from both the training and test sets and randomly sampling the attack records to ensure that those records most difficult to classify were more likely to be included. The sampling mechanism used by Tavallaee et al. assigned records to categories based on the number of the 21 learners (classifiers trained by one of the three samples utilized in the study) that correctly classified the record. The percentage of records included in the NSL-KDD dataset from a particular category is inversely proportional to the percentage of records in the KDD99 dataset of that category. For example, records correctly identified by six to ten of the learners made up 0.07% of the KDD99 dataset, so 99.93% of those records are included in the new dataset. In the process, they reduced the number of records in the training and test sets to a reasonable number, thus allowing use of the full NSL-KDD dataset instead of a sample [54]. The NSL-KDD dataset is beginning to be used in research. Salama et al. [45] tested a hybrid NIAD scheme using a deep belief network for feature reduction and a support vector machine to classify the trace. Wang et al. [60] improved a Distance-based Classification Model, then tested the new system on the NSL-KDD dataset. Iranmanesh et al. [25] demonstrated the efficacy of selecting landmarks using Fuzzy C-Means for Incremental L-Isomap feature reduction by applying the method to the NSL-KDD and other datasets. Lakhina et al. [33] applied a new principal component analysis neural network algorithm to reduce the number of features in the NSL-KDD dataset, resulting in reduced training time and improved accuracy. The NSL-KDD dataset may replace the KDD99 dataset as the baseline in future NIAD research.

2.3.3 Kyoto Dataset.

The Kyoto University Benchmark dataset [49] consists of 3 years (November 2006 through August 2009) of data captured from honey pots, darknet sensors, a mail server, a web crawler, and Windows XP installation. While very carefully constructed and comprehensive, the dataset does not lend itself to evaluation of new NIAD classifiers for several reasons: First, the Kyoto dataset contains only the values of specified features, and lacks the full raw-data packet captures which would allow for implementation of future advances in feature selection and extraction [5]. Second (and more importantly), the accuracy of the labels in the Kyoto dataset is unknown [50]. The dataset consists of captured network traffic automatically labeled using a Symantec Network Security 7160 appliance (discontinued as of December 12, 2008), Clam Antivirus (updated once per hour), and dedicated shell code detection software called *Ashula* [47]. The use of an Intrusion Detection System (IDS) appliance to label records was necessary, as human labeling of that much traffic is impossible, but when automated labeling is used the label accuracy comes into question [42]. Without a certain distinction between normal and intrusion records, any NIAD evaluations based on the Kyoto dataset are subject to error.

2.3.4 Other Datasets.

Several additional datasets of limited usefulness for evaluation of NIAD systems are also publicly available, including the Information Exploration Shootout (IES) [20], NIMS1 [1], and University of Cambridge (UC) [40] datasets. The IES datasets consist of four attack datasets, each of which contains only a single attack type, and a baseline set with no attacks [19]. The IES datasets contain samples from only four attack types, as opposed to the KDD99 dataset which contains data from a total of 22 different types. This lack of variety in attack types may restrict the usefulness of IES as a dataset for NID algorithm validation. The UC datasets are focused on classification of traffic types, and do not contain detailed labels for the attacks [39], which may limit their usefulness in NID algorithm validation.

The NIMS1 datasets are focused on encrypted traffic classification: They do not contain any labeled attacks or intrusion attempts [2], and may not be useful in NID algorithm validation.

2.4 Dataset Characterization

There are two key problems in NI dataset characterization: (1) Evaluation of the accuracy of the labels within the dataset and (2) Evaluation of how well the dataset represents the target traffic. Labeling has been addressed extensively in the literature, but there has been little focus on dataset representation . One study by Joseph et al. [28] proposed a unique validation method based on a community detection algorithm proposed by Schuetz et al. [46]. The algorithm is a very fast method of detecting the communities (groupings) among a network based on the difference between the number of connections between vertices and the number of connections which would exist between those vertices were the connections made randomly. The community detection algorithm results in a reasonable estimation of the communities extant within a network. Joseph et al. use the algorithm proposed by Schuetz et al. to produce a visualization of networks showing their communities. Multiple graphical entities are produced to allow visual analysis and comparison of datasets. Circles represent communities of hosts using particular protocols, where the diameter of the circle is proportional to the number of hosts within the community. Lines between the circles represent connections between hosts in the two communities, and the thickness of the line is proportional to the number of connections between the two communities. See Figure 2.1 for an example visualization, where the communities A, B, and C have 15, 20, and 30 hosts, respectively; and lines AB, BC, and AC consist of 15, 25, and 35 connections, respectively.

Throughout the visual verification study, the supposition is that both dataset validation and verification of the topology and protocol distribution of multiple networks is being accomplished. However, the study does not specify which qualities of a dataset are being examined to determine the validity of each dataset. Examining the topological properties of

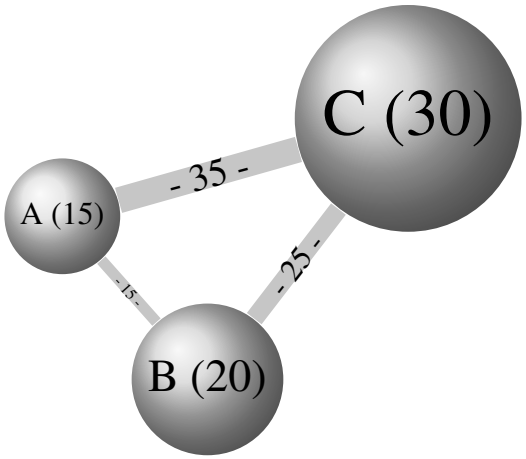

Figure 2.1: Example visualization of a small network consisting of three communities.

a network may be useful for general network traffic analysis, but defining which qualities of the topological properties will be used to validate one network against another is necessary if validation is the goal. The graphical visualization techniques and methods proposed in the visual verification study may be useful for traffic distribution analysis, as the visualization presented in the study is of traffic distribution and common protocol communities. The visual verification approach may also provide a view of how well a dataset consisting of generated traffic represents a particular network, but only in terms of protocol distribution and connections between communities. Joseph et al. acknowledge that their method is only useful to distinguish datasets with flawed topological and protocol distributions, and does not completely verify the quality of datasets. As proposed, the visual verification gives researchers a method of quickly viewing groupings within two networks to determine a rough estimate of their comparability. There is also a significant problem in using this method on collected traffic, as there is no verification of the traffic label accuracy. While the proposed visual verification method could be used as part of a general dataset validation methodology, alone it is unsuitable for quantitative NI dataset validation.

Other researchers have worked on accurately classifying packets into different categories to better understand the character of network traffic. Within the field of computer network traffic classification, there have been many studies on how to categorize network traffic, some of which are collected by Zhang et al. [62]. Focus has ranged from developing software to assist in the comparison of different NIAD classifiers on the same datasets [34] to an evaluation of the utility machine learning algorithms might have in network traffic classification [51]. All of these classification methods could be useful in determining overall patterns and distributions of network traffic. Identification of patterns and distributions would then be useful in developing models for generating background traffic to be used in NI datasets.

The severity of the labeling and representation accuracy problems depends on the type of dataset being created. If properly collected, datasets based on captured raw traffic can be assumed to be representative of the target network traffic [44]. However, the labels of a captured dataset are suspect, and lead to uncertainty for any studies using that dataset. Generated datasets, on the other hand, have accurate labels. However, there is currently no identified way to ensure that a given generated dataset is sufficiently representative of the target network traffic. As to hybrid methods of dataset creation, where there is a mix of captured live traffic and generated traffic, it is unknown whether the labeling and representation problems would be aggravated or diminished. One study by Mahoney and Chan [37] presented results suggesting that simulation artifacts in generated traffic can be eliminated by careful mixing of raw and generated traffic, but more work remains to generalize those results. As summarized in Table 2.1, this review has not discovered a method of dataset creation which results in both representative traffic and accurate labels.

Much research has been done on characterizing the IDEVAL [36] and derivative KDD99 [53] datasets. Critiques of the problems inherent in those datasets serve to highlight problems with generated datasets in general, and the methods used in their generation.

Table 2.1: Certainty of characterization of dataset creation methods

	Certainty of Characterization	
Dataset type	*Labels*	*Representative*
Captured	Uncertain	Certain
Generated	Certain	Uncertain
Hybrid	Uncertain	Uncertain

McHugh's [38] analysis of IDEVAL resulted in recommendations including: The definition of metrics to measure performance, a calibrated and validated field-wide generated dataset or dataset generator, studies of false alarm rates in generated and collected datasets (to establish a relationship), research into creation of a field-wide set of network attack taxonomies, and a field-wide sharable source of recent attacks for use in research. Mahoney and Chan [37] performed a thorough examination of possible simulation artifacts within the IDEVAL datasets, and concluded that mixing real traffic with simulated traffic can remove those artifacts leading to a better evaluation of NIDS capabilities. Multiple studies of IDEVAL [16, 57] brought forth few new significant problems which generalize to simulated dataset generation not already noted, although they did conclude IDEVAL is still useful as a baseline for use in research.

NIAD research has been going on for many years. In spite of progress made in areas of NIAD research, including classification [30], anomaly detection [7], signature based detection [58], preprocessing [12], feature selection [8], and other topics; minimal work has been done on improving the pedigree of datasets used to validate NIAD classifiers. Many studies were performed with datasets local to the organization performing the research, or

originating from packet captures restricted from general use for legal reasons. With the different datasets used in performance evaluations over the years, it is difficult to accurately compare the results of one to another.

When characterizing datasets, there is implicit in the characterization the ability to compare one dataset to another. With the ability to compare one dataset to another, comes the concept of the difference between datasets and a quantifiable measure of that difference: the difference measure.

2.5 Difference Measures

There is a need to understand the current state of how distance measures are used in the field of NI. Knowledge of how the distance and similarity measures are used is insufficient, however, without knowledge of how well those measures used are identified by name or formula (or both) within the recently published papers. To provide a thorough overview, a recursive automated search is executed for the terms "network intrusion" in the title, abstract, or keywords using Google Scholar. The number of papers returned is limited by restricting the results to the first 100 returned papers, the depth of the reverse citation lookup to three, and the year published to between 2008 and 2012 (inclusive). The results provide the source for a random sample of the field.

There were 2,235 results returned by the search procedure. Due to time constraints a two pass method was used to eliminate papers. In the first pass, suitability for inclusion within the study is determined. Each paper is examined by hand and deemed unsuitable if it is a duplicate of a paper already examined, unrelated to the NID field, or unavailable electronically. Most of the papers are eliminated by that pass, leaving only 567. In the second pass, the remaining 567 papers are more closely examined to eliminate reviews, surveys, overviews, or examinations of the state of the art. The remaining 536 papers represent the most recent published papers in the field of NI.

As a full examination of all 536 papers is not possible within the time available, approximately 20% of the papers are randomly selected for inclusion in this sample. Each paper is examined to check for the following criteria.

1. The paper must contain results from classifying at least one NI dataset with the purpose of identifying attacks.

2. The paper must be available in English, or have an identification and formulation of distance measure use understandable regardless of the language.

The five papers that do not meet the two criteria above the number are discarded, leaving 100 papers. Each of those 100 papers is then examined to determine how (and if) the distance measures used within were named and formulated. For the full list of papers included in the survey, see Appendix A.

2.5.1 Evaluation of Measure Specificity.

As there sometimes exist differences in the terminology used to describe measures used, or between the formulation of each measure and those used in other works, this research uses the Encyclopedia of Distances [15] as the standard listing of measures. Any names and formulas found in the sample are compared to those within the Encyclopedia. It is useful to categorize the papers based upon explicit naming and formulation of distance measures used within the works, and the existence of those names and formulas within the Encyclopedia of Distances, effectively separating them into the categories of *Given* and *Not given*. However, it is apparent that some authors have developed new distance measures, or new names for existing distance measures, in response to the specific needs of the research being conducted. When the newly specified names and formulas are used, and are not identifiable in the Encyclopedia of Distances, an additional category is necessary: the *Novel* category.

To quantify how well the name and formulation of each distance measure matched the standard, a single set of categories are applied to the names and formulas used within each work, and a single definition of each category is used.

Among the works sampled for this review 60 of the papers did not provide a measure name, and 68 of the papers did not provide an explicit formulation. The sample taken indicates that most of the field does not provide names and formulas for the distance measures used within their research, and thus makes replication difficult. Ensuring clarity when specifying which and how each distance measure is used is critical to repeatability of published works. Vague descriptions and assumed measures lead to confusion at best, and incorrect implementations when attempting to duplicate experiments (thus incomparable results) at worst.

It is evident many different distance measures are used by researchers, but unfortunate how rarely they specify precisely which measures are used. One possible reason for the lack of specificity is the difference in terminology which exists among fields [15]. In example, if an author came to the AD field from the field of Ecology, or were familiar with the use of similarities in that field, they may use the term, "niche overlap similarity," [43] instead of, "cosine similarity," resulting in confusion to those unfamiliar with published works in the field of Ecology. An examination of the methods used by authors that do explicitly provide the name and formula of the measure(s) used, along with how clearly they explain the application of the measure and the phase in which it was used, is undertaken below.

2.5.2 Evaluation of Measure Types Used.

It is useful to understand which types of measures are being used in the field of NIAD, and which are not. When specifying the type of distance measure used, the names of types defined within the Encyclopedia of Distances are used: power (p, r)-distance, distances on distribution laws, and distances on strings and permutations. Disregarding the works without specified distance measures, there are 40 works within which distance measures are at least named, and three types of distances identified within the sampled works.

The first type of distance, the power (p, r)-distance, is shown in Equation (2.1). All instances of power (p, r)-distance observed within the sample used either the formula in Equation (2.1), or one that is mathematically equivalent.

The second and third types of distance used within the sample do not have a specific formula, as there is more variation among the implementations. The distribution of distance measures within the 40 papers is listed in Table 2.2.

Table 2.2: Frequency of difference measure types used within sampled works.

Measure type	Number of papers
Power (p, r)-distance	22
Distances on distribution laws	15
Distances on strings and permutations	3

The majority of the distance measures used within the sample are based on the power (p, r)-distance and distances on distribution laws. Of the 22 uses of the power (p, r)-distance, 40% used Euclidean distance ($p = r = 2$). The sample shows that there is little exploration of the possible use of other measures within the field, as most of the measures are based upon the power (p, r)-distance or distances on distribution laws.

2.5.3 Distance Measure Categories.

In examining the types of distance measures used within the NI field, it is useful to consider distance measures as part of distinct families or categories. The families selected for this work are among those enumerated in, "Encyclopedia of Distances," by Deza and Deza [15], where it is noted that the selection of a similarity index or distance measure is dependent in large part on the data. As there is no definitive taxonomy within the NID field, the measures and indexes examined will be ordered by their relationship to families of measures. As there are many published works in the NID field which do not identify

and specify the distance measures used, the focus in this section is to examine those papers which do both. This section provides examples of works in which the authors identify or formulate the measures used in a manner conducive to repetition or extension of their work. In particular, specificity when providing the formula for a distance or similarity measure is useful when repeating or extending works, and should be the standard rather than the exception. Unless otherwise noted below, the papers examined within this section provide good examples allowing duplication of the performed experiments by future researchers.

During this review it became evident that the three measure families most commonly used are those related to power (p, r)-distance, distances on distribution laws, and correlation similarities and distances. First, and most common, power (p, r) distance has already been defined in Section 2.5.2 and formulated in Equation (2.1). Second in popularity are distances on distribution laws, measures that apply to probability distributions over variables with the same range. Finally, correlation similarities and distances are measures that attempt to characterize the correlation between two datasets, and treat that as a measure of similarity or distance, rather than using the probability distributions or magnitude of vectors.

A large number of works within the NI field use power (p, r)-distance related distance measures. In particular many use Euclidean distance, which is an expression of Equation (2.1) where $p = r = 2$. To provide some intuition about what different p and r values would mean when measuring the distance between two points, Figure 2.2 shows Voronoi diagrams [3] constructed using the power (p, r)-distance while varying the value of p and r.

After the measures related to power (p, r)-distance, those related to distances on distribution laws, or probability, were the next most common within the sample of the field (see Subsection 2.5.1) and literature review. While there is almost always an aspect of probabilistic behavior to any data collection and analysis, in this review the focus is on those studies which specifically used probability in one of the NID processes identified in Section 2.2.

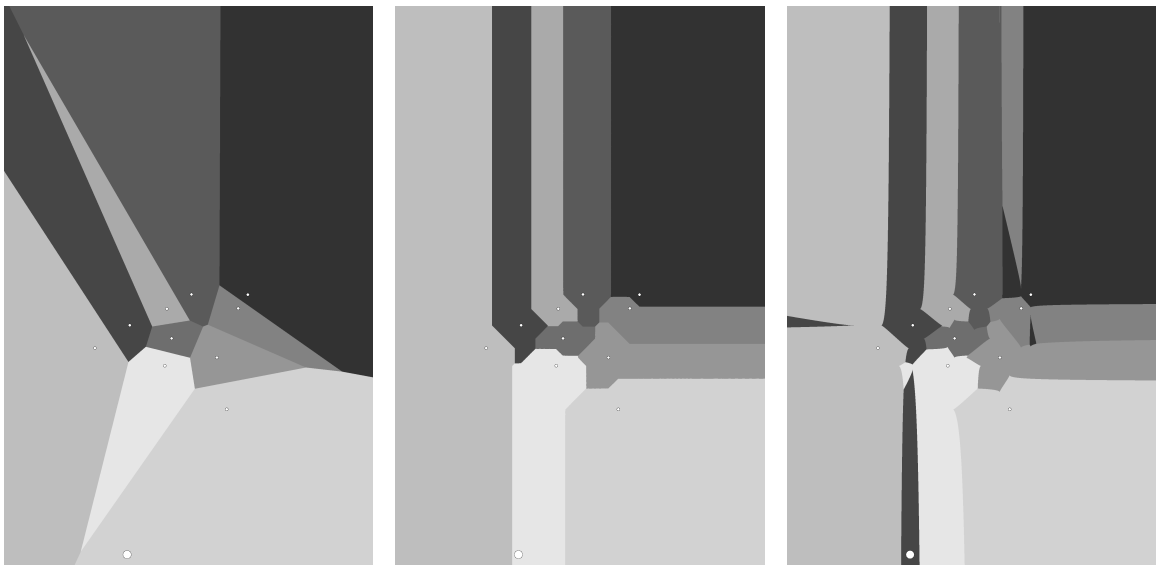

Figure 2.2: Voronoi diagrams using power (p, r)-distance with $p = r = \langle 2, 1, 0.75 \rangle$ (from left to right).

One study in particular uses probability-related measures during feature selection. Hancock and Lamont [23] develop a, "Multi agent system for network attack classification using flow-based intrusion detection," using the Bhattacharyya coefficient to rank features according to the ability of each feature to distinguish one class from the others. The Bhattacharyya coefficient is shown in Equation (2.3), where P_1 and P_2 are probability distributions over the domain X, $p_1(x)$ is the probability of x occurring in P_1, and $p_2(x)$ is the probability of x occurring in P_2.

$$\rho\left(P_1, P_2\right) = \sum_{x \in X} \sqrt{p_1(x) p_2(x)} \tag{2.3}$$

The three features with the largest overlap (largest $\rho\left(P_1, P_2\right)$ value) are selected after rejecting any feature which is strongly correlated to a higher ranked feature to reduce redundancy among selected features. The feature selection is part of their second design iteration while pursuing the goal of an effective multi-agent NID system using reputation.

Hancock and Lamont identify the distance measure by name, but do not specify a formula with which to verify the work.

The least common measures used are those based on correlation similarities and distances. Zhao et al. [63] use a single distance measure which incorporates one of three correlation coefficients to detect stepping-stone attacks, where one computer is used by the attacker to reach another in, "Correlating TCP/IP Interactive Sessions with Correlation Coefficient to Detect Stepping-Stone Intrusion."

The first correlation is the Spearman ρ rank correlation. Equation (2.4) gives the Spearman ρ rank correlation where X_r and Y_r contain the rankings of discrete variables X and Y, x_i and y_i contain the i^{th} rank in X_r and Y_r (respectively), X_r and Y_r have the same number of elements, and n is the number of elements in X_r.

$$\rho(X_r, Y_r) = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2-1)} \tag{2.4}$$

The second correlation coefficient is similar to the Kendall τ rank correlation. To properly define the Kendall τ rank correlation a preliminary definition is required. The *sign*, or *signum*, function is defined in Equation (2.5).

$$\operatorname{sign}(x) = \begin{cases} -1, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \\ 1, & \text{if } x > 0 \end{cases} \tag{2.5}$$

The Kendall τ rank correlation can then be defined using the *sign* function as in Equation (2.6), where the *sign* function is used to calculate the number of discordant pairs of ranks subtracted from the concordant pairs of ranks.

$$\tau(X_r, Y_r) = \frac{2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\operatorname{sign}(x_i - x_j)\cdot\operatorname{sign}(y_i - y_j)}{n(n-1)} \tag{2.6}$$

The authors use an equivalent formulation, but to properly define the equivalent formula a preliminary definition of the *equals* function is provided, as shown in Equation (2.7).

$$\text{equals}(x, y) = \begin{cases} 1, & x = y \\ 0, & \text{otherwise} \end{cases} \tag{2.7}$$

The authors formulation of τ can then be defined using the *equals* and *sign* functions, and is shown in Equation (2.8).

$$\tau(X_r, Y_r) = \frac{4 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{equals}(\text{sign}(x_i - x_j), \text{sign}(y_i - y_j))}{n(n-1)} - 1 \tag{2.8}$$

The third correlation coefficient is the Pearson product-moment correlation linear coefficient, or r. The formula for r is given in Equation (2.9), where $\bar{X}$ is the mean of the discrete variable X.

$$r(X, Y) = \frac{\sum_{i=1}^{n} \left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right)}{\sqrt{\sum_{j=1}^{n} \left(x_i - \bar{X}\right)^2 \sum_{j=1}^{n} \left(y_i - \bar{Y}\right)^2}} \tag{2.9}$$

The authors use an equivalent formulation, defined in Equation (2.10).

$$r(X, Y) = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \sqrt{n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \tag{2.10}$$

Each measure is applied to the two traffic streams, and each result is subtracted from one, to calculate the minimum distance between the two streams ($\sigma(X, Y)$), as shown in Equation (2.11).

$$\sigma(X, Y) = \min(1 - \rho(X_r, Y_r), 1 - \tau'(X_r, Y_r), 1 - r(X, Y)) \tag{2.11}$$

If $\sigma(X, Y)$ is less than a threshold set by the researcher, then the compared pair is relayed traffic. The use of multiple correlation coefficients within a single distance measure

is a good example of using multiple measures to detect similarities a single measure might not catch.

The predominant use of power (p, r)-distance and distances on distribution laws provides guidance on areas which have been explored in the current body of research. Also, the clear focus within the research on one or two measures per study is interesting, as it indicates further study of a multitude of measures applied to a single problem is, effectively, an open area.

Every experiment which utilizes AD in the NIAD field uses distance measures, most without much thought as to which distance measure would be most appropriate. Current research in the NIAD field is focused on the use of NIDS, rather than examining and characterizing NI datasets and the networks from which they are derived. However, the ways in which some distance and similarity measures are used can still be helpful in the comparison of datasets, as classifiers are, in some sense, distance measures: The accuracy of a specific classifier on a dataset, when compared with its accuracy on another, provides some knowledge of the difference between the two datasets.

The analysis and characterization of datasets which happens during the preprocessing, feature generation, and feature selection phases of NIDS is exactly what is needed when determining how to compare datasets. Using the knowledge, techniques and tools available in the NI field today, it should be possible to develop a methods to compare and characterize datasets without having to evaluate them using classifiers and feature selectors.

III. Methodology

The *performance*, or ability to successfully detect attacks, of a Network Intrusion Anomaly Detection (NIAD) classifier depends heavily on the characteristics of the network. The characteristics include the types of network traffic, frequency of attacks, and types of attacks on that network. Because classifier performance depends on those factors, and because those factors and classifier performance vary from network to network, there may be a correlation between classifier performance on two networks and the differences between those networks (provided the proper measure of difference can be found). In the NIAD field networks are commonly examined through logs of network traffic that are converted into datasets. As the focus of this work is on networks which are examined through datasets, within this work *network* will be used interchangeably with *dataset*. The hypothesis to be tested follows:

Hypothesis. *The difference between two datasets, when calculated using the proper difference measure, is correlated to the change in classifier performance between the two datasets.*

To detect a correlation between the performance of a classifier on two datasets and the difference between the two datasets, there must be precise definitions of classifier *performance* and the *difference* between networks. While usually distance is used to denote a measured value representing the dissimilarity between two entities, distance is often interpreted as requiring non-negativity. For that reason, difference is used throughout this work to refer to the calculated measure of dissimilarity between two entities.

To measure classifier performance, three separate methods will be used. The first is the classifier *accuracy* (A), which gives the overall ratio of correct classifications to number of

samples. To calculate the accuracy of the classifier Equation (3.1) may be used:

$$A = \frac{P_t + N_t}{P_a + N_a} \tag{3.1}$$

where P_a is the number of Actual Positives (AP) (total anomalous samples), P_t is the number of True Positives (TP) (correct anomalous classifications), N_a is the number of Actual Negatives (AN) (total normal samples), and N_t the number of True Negatives (TN) (correct normal classifications).

The second measure used is the *true positive rate* or *sensitivity* (R_{P_t}), which gives the proportion of correctly classified anomalous samples to the total number of anomalous samples. R_{P_t} can be calculated using the Equation (3.2):

$$R_{P_t} = \frac{P_t}{P_a} \tag{3.2}$$

The third measure used is the *false positive rate* (R_{P_f}), which provides the proportion of incorrectly classified normal samples to the total number of normal samples. To calculate R_{P_f}, first the number of false positives must be calculated using Equation (3.3):

$$P_f = N_a - N_t \tag{3.3}$$

Once P_f is known, then R_{P_f} can be calculated using Equation (3.4):

$$R_{P_f} = \frac{P_f}{N_a} \tag{3.4}$$

With a clear definition of classifier *performance*, the focus can shift to defining what is meant by the *difference* when referring to datasets. Difference cannot initially be defined as clearly as performance, as a primary goal of this research is identifying one or more difference measures that properly characterize the difference between two networks and correlates well with the difference between classifier performance on the two networks. However, there are some characteristics of the difference between datasets that can be defined clearly. The following characteristics are useful qualities for a difference measure in this research.

- The difference measure must be a scalar, to provide a summary of dissimilarity which is also usable when developing a predictive model.
- The difference measure should provide both magnitude and sign, to adequately correspond to the sign of the change in accuracy.

In addition to defining a standard measure of difference between networks, identifying a difference measure which satisfies the above characteristics may be useful in predicting the performance of many kinds of classifiers on one network, based on the difference between that network and another.

3.1 Approach

To determine a suitable difference measure, identifying those elements of networks which have the greatest effect on classifier performance is key. NIAD classifiers depend on the separability of anomalous traffic, thus examining the ways traffic can be similar provides a useful starting point for difference measure selection. In general, the Anomaly Detection (AD) methods used in measuring differences between samples can be separated into two broad categories: spatial and probabilistic.

Spatial methods use the features of two samples as coordinates within the sample space, and compare the distance between two or more sets of samples to determine the difference between the samples. Those methods are then generalized to more than two samples, and used to measure the difference between two datasets. Probabilistic methods use the distribution of values within the features of a dataset, instead of the values themselves, to determine the overall probability distribution of the samples within the dataset. Measuring the difference between two datasets is accomplished by calculating the difference between the entropy of each dataset, using the probability distribution to calculate the entropy. As both methods can be useful in measuring difference, measures from both categories are used to measure difference in this work.

In developing experiments to test the Hypothesis 3, it is first useful to provide a structure that can be used to perform experiments. To do so, a System Under Test (SUT) is developed which represents the possible inputs, processes, and outputs involved in testing Hypothesis 3.

3.2 System Boundaries

The System Under Test (SUT) is the Classifier Accuracy Prediction System (CAPS) shown in Figure 3.1. The CAPS consists of an Anomaly Detection (AD) classifier, difference measure, correlation function to indicate if change in performance has a linear or rank-based correlation with the difference between the datasets, a linear model generator, and a predictive error calculator. Determination of how well a classifier performed on a given dataset is evaluated relative to how well it performed to another dataset. No attempt is made to determine which classifier is better, and feature generation and selection are not considered part of the test.

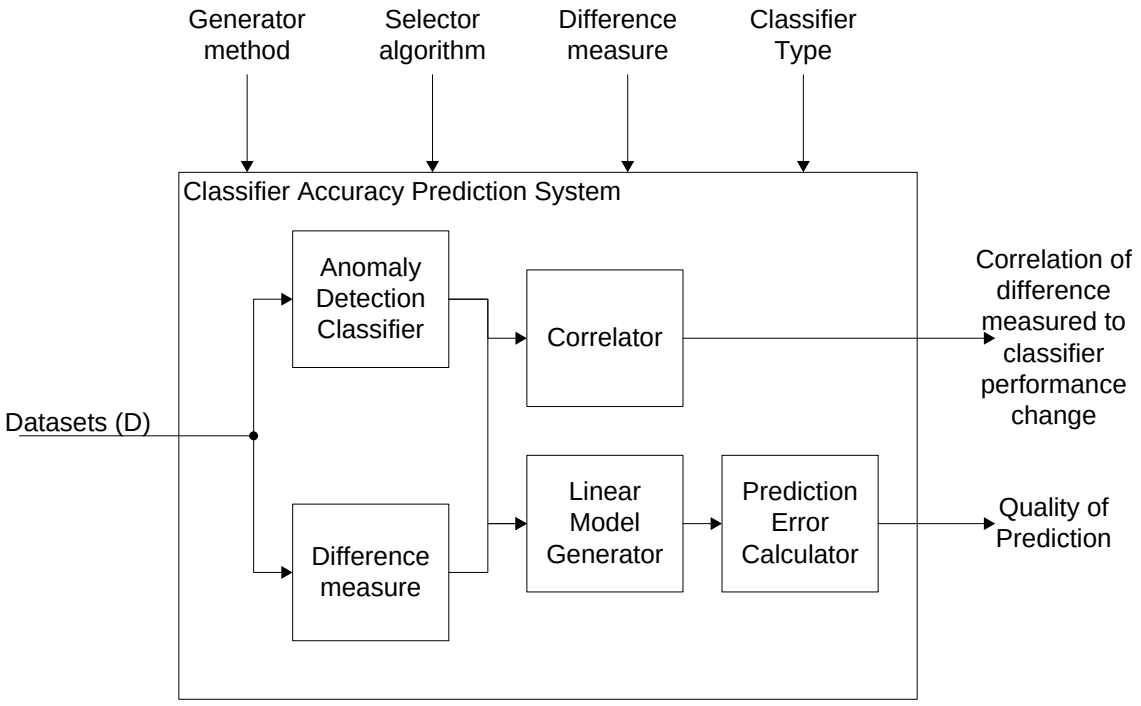

Figure 3.1: The System Under Test - CAPS

The Component Under Test (CUT) is the difference measure. For identical datasets, upon which any classifier should perform identically, the difference measured should be statistically zero. For different datasets, the difference measured should be statistically non-zero.

3.3 System Services

The CAPS provides two services. The first is a set of values indicating how the difference measured among each pair of datasets correlates to the change in classifier performance between each corresponding pair of datasets. The difference is calculated using all difference measures. The second is the set of linear model prediction error values. To be considered successful, the difference measured between each pair of datasets must correlate consistently with classifier performance change between the same pair of datasets, and the linear models must be able to predict the changes in classifier accuracy based only on the differences measured between the datasets.

The CAPS takes as input a set of datasets (D). Each dataset is classified using an AD classifier. The difference ($\delta(i, j)$) between every pair of datasets such that $\left\{d_i, d_j \in D \,|i \neq j\right\}$ is calculated. The change ($\varsigma(i, j)$) in classifier performance between each pair of datasets is also calculated. The correlations of the set of $\delta(i, j)$ and $\varsigma(i, j)$ are calculated. The potential value of each correlation coefficient ranges from negative one to one. A zero means no significant correlation was found between $\delta(i, j)$ and $\varsigma(i, j)$, whereas a negative one or one means $\delta(i, j)$ and $\varsigma(i, j)$ are perfectly correlated.

3.4 Workload

The workload consists of the datasets being compared by the CAPS, and the features included in each dataset. The features of the dataset are considered part of the workload parameters because they contain the values being considered when differences are calculated, therefore the correlation is based on these features.

3.4.1 Datasets.

In Network Intrusion Anomaly Detection (NIAD) research, there are two primary Network Intrusion (NI) datasets which have been used in the examination of AD classifiers: the 1999 Knowledge Discovery and Data Mining Tools Competition (KDD99) dataset [53] and the NSL-KDD [55] dataset. While those are ideal datasets to evaluate candidate difference measures once the candidates have been identified, time-constraints demand the use of less-complex datasets in the initial experiment.

The datasets used are a set of 38 2-D Synthetic Datasets [27], developed by Ji and Dasgupta [26]. They are designed to be used in AD classifier evaluation, and represent a variety of geometric shapes in different sizes as demonstrated by the visualization of one dataset in Figure 3.2. As the training datasets contain only self or non-self samples (depending on the perspective of the tester), only the test sets were used for this experiment.

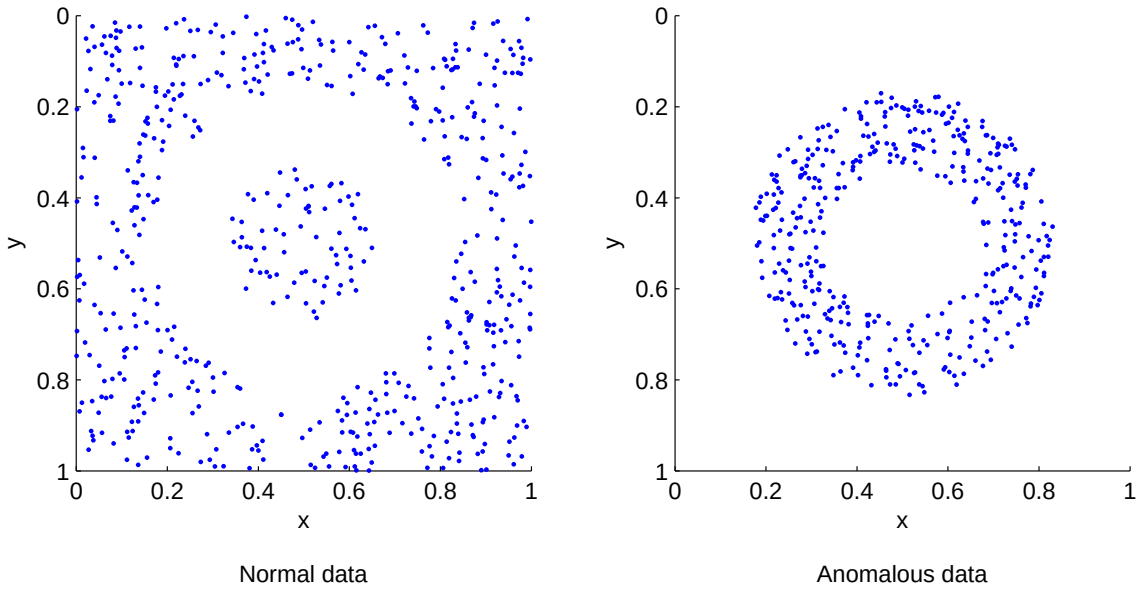

Figure 3.2: The normal and anomalous data points in *Ring_midneg_test* dataset. This is one of 38 synthetic datasets developed to test the abilities of AD classifiers.

As a result of using synthetic datasets of this type, there are limitations with regards to the claims that can be made based on the experimental results. First, the results may not be applicable to all datasets used within the NIAD field, but only to those which use datasets similar to those used in this study. Second, by using datasets designed with geometric shapes embedded within the data, there is an assumption made that datasets of interest will have shapes, and that the shapes will define the separation between anomalous and normal (non-self and self). That assumption may not be true for datasets commonly used in the NIAD field. Third, and finally, the synthetic datasets are two dimensional, whereas NI datasets can have tens or hundreds of dimensions. The results obtained here may not be applicable to datasets with higher dimensions.

Each of the datasets used consists of 1000 samples with two features: x and y coordinates on a plane. Although it is common in the classification field to label anomalous and normal samples with a one and zero (respectively), the authors of these datasets labeled the anomalous and normal samples as zero and two (respectively).

3.5 Performance Metrics

The two primary performance metrics are the correlation coefficients used to determine the correlation between between the change in classifier performance and difference between datasets, and the quality of predictions by the linear models. The experiment is considered successful if there is consistent non-zero correlation between the dataset differences and corresponding changes in classifier performance, *and* if the prediction quality is non-zero.

The first performance metric is composed of two correlations. The first correlation is the Spearman rank correlation coefficient (ρ). The calculation of ρ requires the ranks of the data, not the values. The sign of ρ indicates the direction of association between the compared variables. The magnitude of ρ indicates whether the value of the response variables (change in A, R_{P_t}, and R_{P_f}) are related monotonically to the predictor variable (difference between datasets). Given that x and y contain the rankings of the predictor and

response variables (respectively), I contains the total number of values, and n contains the total number of ranks, ρ can be calculated using Equation (3.5):

$$\rho = 1 - \frac{6}{n(n^2 - 1)} \sum_{i \in I} (x_i - y_i)^2 \tag{3.5}$$

The second correlation is the Pearson product-moment correlation coefficient (r). The calculation of r indicates the linear dependence of the two variables. Given that x and y vectors that contain the values of the predictor and response variables (respectively), and n contains the total number of values, r can be calculated using Equation (3.6):

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{3.6}$$

The second performance metric is the quality of the linear model predictions. The quality of prediction will be measured using the error between predicted and actual accuracies, or *residuals*, by examining the distribution of the residuals with respect to the distribution of the actual classifier performance values. Any ability to predict the actual performance values will be considered success.

In addition to the two checks for correlation and prediction quality, the results will be evaluated visually to ensure the results make sense. The first visual check is a series of scatter plots, using the predictor variable as the x-coordinate and the response variable as the y-coordinate.

The second visual check is a series of Quantile-Quantile (QQ) plots. A QQ plot is a probability plot which compares two probability distributions (histograms) using the same bins for each. The QQ plot is often used to compare a sample to a standard distribution, but instead it will be used to compared the distributions of differences and performance changes. Ideally, if they are from the same distribution they will align linearly along the diagonal.

3.6 System Parameters

The following parameters affect the correlation between performance and difference.

3.6.1 Difference Measure.

The difference measured between two datasets ($\delta(i, j)$) should determine how far apart two datasets are with respect to those characteristics which affect classifier performance. It affects the correlation significantly, and six different measures are used to quantify the difference between datasets.

The first five difference measures use entropy, as entropy gives an estimate of the predictability of the values in a dataset. The ability to calculate entropy requires the probability distribution of each feature (or each set of features), therefore the probability distribution must be calculated first. Calculating the probability distribution can be problematic in some datasets, because sometimes no value is repeated within a feature. As a result of that limitation, the probability distribution is estimated as follows. First, a reasonable bin-size is calculated using the Freedman-Diaconis rule [17].

$$h = 2\text{IQR}(X)n^{-\frac{1}{3}} \tag{3.7}$$

The function IQR(X) returns the interquartile range of the data in a feature vector X, and n is the number of observations in X. Once the bin-size (h) is known, the set of bin edges e are calculated:

$$\begin{aligned} l &= \min(X) \\ u &= \max(X) \\ b &= \left\lceil \frac{u-l}{h} \right\rceil \\ e &= \langle -\infty, l+h, l+2h, \ldots, l+(b-1)h, \infty \rangle \end{aligned} \tag{3.8}$$

Those calculations are performed for each feature. Once the edges are known, the values of each feature are sorted into bins while keeping count per bin. When complete the researcher is left with an estimate of the probability distribution of all features ($p(F)$). With $p(F)$, the entropy of each feature can be calculated using the standard formula for entropy:

$$H(X) = -\sum_{x \in X} p(x) \log_2(p(x)) \tag{3.9}$$

where X is a feature in F (the set of all features).

The first difference measure is the difference between the sum of the entropy of each feature in two datasets. It is identified as δ_S, and is calculated between two datasets d_i and d_j shown in Equation (3.10):

$$\delta_S = \sum_{X \in F_i} H(X) - \sum_{Y \in F_j} H(Y) \tag{3.10}$$

where F_i and F_j are the set of features in datasets d_i and d_j, respectively. The second difference measure is the absolute value of Equation (3.10). It is identified as $\delta_{|S|}$, and is shown in Equation (3.11).

$$\delta_{|S|} = |\delta_S| \tag{3.11}$$

The third difference measure is the sum of the absolute value of difference between the entropy of each feature. It is identified as $\delta_{|F|}$, and is calculated between two datasets d_i and d_j as follows:

$$\delta_{|F|} = \sum_{X_i \in F_i, X_j \in F_j} \left| H\left(X_i\right) - H\left(Y_j\right) \right| \tag{3.12}$$

where F_i and F_j are the full set of features within datasets d_i and d_j, respectively.

The fourth difference measure using entropy is the difference between the joint entropy ($H(X_s, X_t)$) of two datasets. The joint entropy of a dataset with two features is calculated using the joint probability distribution ($p(X, Y)$) of the features within the dataset:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2(p(x, y)) \tag{3.13}$$

where X and Y are the two features within a dataset. A joint probability distribution can be calculated for more than two features:

$$H(X_1, \cdots, X_t) = -\sum_{x_1 \in X_1} \cdots \sum_{x_t \in X_t} p(x_1, \cdots, x_t) \log_2(p(x_1, \cdots, x_t)) \tag{3.14}$$

where X_t is the t^{th} feature in the dataset. This formulation is problematic in those datasets where all observations are unique in value, as it will result in maximal entropy, thus providing

zero differentiation between datasets. To prevent the result of maximal entropy the joint entropy is calculated in a different way.

Instead of calculating the joint entropy on the values directly, the probability distribution of each feature is first calculated. A new set of features (F') is formed that has the same number of observations and features as the original dataset. For each feature in F, a probability distribution is calculated using Equations (3.7) and (3.8). Each bin is given a number starting with one for the bin containing the lowest value, incrementing the value by 1 for each bin containing successively higher values. Once the edges and bin numbers are known, the values of each feature are sorted into bins while keeping track of which observations go into each bin. Each observation of F' is then assigned as a value the bin number to which the corresponding observation of F was assigned. The new feature set, F', is then used to calculate the joint probability distribution of all the features. The joint probability distribution is then used to calculate the joint entropy of the dataset, and finally to calculate the difference between the joint entropy of two datasets. It is identified as δ_J, and calculated as follows:

$$\delta_J = H\left(X'_1, \cdots, X'_t\right) - H\left(Y'_1, \cdots, Y'_t\right) \tag{3.15}$$

where X'_t and Y'_t are the t^{th} features in the feature sets F'_i and F'_j, respectively. The fifth difference measure is the absolute value of Equation (3.15). It is identified as $\delta_{|J|}$, and is shown in Equation (3.16).

$$\delta_{|J|} = |\delta_J| \tag{3.16}$$

The last distance measure is Procrustes distance, and has the limitation that all compared datasets must have the same number of samples. The dissimilarity measure provided by Procrustes Analysis is a summary statistic of the Euclidean distance (Equation (2.1) where $p = r = 2$) between each corresponding point in two datasets after the second dataset is translated, reflected, rotated, and scaled to optimally superimpose it onto the first dataset.

3.6.2 Classifiers.

To determine correlation between the difference value and actual classifier performance, a classifier must be applied to the dataset. Four classifiers were chosen as representative of the various types of classifiers.

First, to represent a robust type of linear classifier the Support Vector Machine (SVM) classifier is used. The SVM classifier first maps the data into a higher dimensional space, and then attempts to find a separating hyperplane which maximizes the separation between the hyperplane and those data points closest to the hyperplane (the *support vectors*) within the higher dimensional space.

Second, to represent the discriminant analysis type of classifier a Quadratic Discriminant Analysis (QDA) classifier is used. The QDA classifier does not assume the covariance of each class is identical, unlike Linear Discriminant Analysis (LDA).

Third, to represent the Bayesian type of classifier (one which uses Bayesian probability theory to calculate posterior probabilities based on current knowledge of priors) the Naïve Bayes (NB) classifier is used. The NB classifier with a kernel distribution is used, as it is appropriate for features which may have a probability distribution with multiple peaks or high skewness.

Fourth, to represent the tree type of classifier, Binary Classification Tree (BCT) classifier is used. The BCT classifier is a simple implementation of the tree type which uses binary splits for classification. It can produce arbitrary non-linear boundaries for classifying the features.

Stratified K-fold cross-validation is used to reduce classifier over-fitting while retaining the proportion of normal to anomalous data in the dataset being partitioned. Each fold will thus have approximately the same proportion of normal and anomalous traffic as that of the dataset from which is was derived.

3.7 Factors

Table 3.1 shows the factors varied during the experiment. Parameters not shown are held constant while the factors below are varied.

Table 3.1: Factors and levels.

Level	Classifier	Difference Measure	Classifier Performance
1	NB	δ_J	A
2	SVM	δ_S	R_{P_t}
3	BCT	$\delta_{\|F\|}$	R_{P_f}
4	QDA	$\delta_{\|J\|}$	
5		δ_P	
6		$\delta_{\|S\|}$	

Classifier

The classifier is varied and applied to each dataset. In this way, the correlation between the performance of multiple classifiers and difference measures can be examined. The following classifiers are used: SVM, QDA, NB, and BCT. As the correlation between changes in classifier performance and difference in datasets is the primary metric, having a wide range of classifiers ensures broad coverage.

Difference measure

The method of quantifying differences between datasets is varied and applied to each unique pairing of datasets. Six methods are computed: $\delta_J, \delta_S, \delta_{|F|}, \delta_{|J|}, \delta_P$, and $\delta_{|S|}$.

Classifier performance

The method of measuring classifier performance is varied and calculated for each classifier and dataset. Three methods are computed: A, R_{P_t}, and R_{P_f}.

As the interactions between the various factors are not known, a full-factorial design is used. The three factors involved in this experiment are the classifier type (SVM, QDA, NB, and BCT), classifier performance measure (A, R_{P_t}, and R_{P_f}), and difference measure ($\delta_J, \delta_S, \delta_{|F|}, \delta_{|J|}, \delta_P$, and $\delta_{|S|}$). The three factors have four, six, and three levels (respectively), so a total of $4 \times 6 \times 3 = 72$ experiments will be required.

3.8 Result Analysis

As there are 38 datasets, and the differences between each possible combination of two datasets will be calculated, there will be 703 sets of results for each classifier and 703 results for each difference measure. This follows from the formula used to determine the number of combinations which can be selected from a set of objects:

$$\binom{n}{r} = \frac{n!}{(n-r)!r!} \tag{3.17}$$

where n is the total number of datasets, and r is the number of datasets to select. Thus the number of ways to select two datasets from 38 is as follows:

$$\binom{38}{2} = \frac{38!}{(38-2)!2!} = \frac{38!}{36!2!} = \frac{38 \cdot 37}{2} = 19 \cdot 37 = 703 \tag{3.18}$$

3.8.1 Correlation Between Performance and Difference Measures.

Once classification is complete the change in accuracy, True Positive Rate (TPR), and False Positive Rate (FPR) must be calculated for each unique combination of datasets. As the mean True Positives (TP) (P_t), False Positives (FP) (P_f), Actual Positives (AP) (P_a), and Actual Negatives (AN) (N_a) values for each dataset are already available from the results of the classification runs using stratified K-fold cross validation, it remains to determine the classifier performance for each selected performance measure. The formulas to classify the accuracy (A), TPR (R_{P_t}), and FPR (R_{P_f}) for any pair of datasets (d_i and d_j) follow:

$$
\begin{aligned}
R_{P_t} &= \frac{P_t}{P_a} \\
R_{P_f} &= \frac{P_f}{N_a} \\
N_t &= N_a - P_f \\
A &= \frac{P_t + N_t}{P_a + N_a}
\end{aligned}
\tag{3.19}
$$

For each unique combination of datasets calculate the difference using each of the six difference measures. Next, calculate the change in performance measure for each of the three performance measures. Once complete each distinct pair of datasets have the following associated values.

- Difference between the sum of entropies of each feature (δ_S, see Equation (3.10)).
- Absolute value of the difference between the sum of entropies of each feature ($\delta_{|S|}$, see Equation (3.11)).
- Difference between the joint entropy of each dataset (δ_J, see Equation (3.16)).
- Absolute value of the difference between the joint entropy of each dataset ($\delta_{|J|}$, see Equation (3.16)).
- Sum of the difference between the entropies of each feature ($\delta_{|F|}$, see Equation (3.12)).
- Procrustes distance (δ_P).
- Difference between QDA classifier accuracy.
- Difference between QDA classifier TPR.
- Difference between QDA classifier FPR.
- Difference between NB classifier accuracy.
- Difference between NB classifier TPR.

- Difference between NB classifier FPR.
- Difference between BCT classifier accuracy.
- Difference between BCT classifier TPR.
- Difference between BCT classifier FPR.
- Difference between SVM classifier accuracy.
- Difference between SVM classifier TPR.
- Difference between SVM classifier FPR.

Each unique combination of classifier, performance measure, and difference measure are then correlated to determine if there is a statistically significant correlation of difference measure to performance measure.

3.8.2 Classifier Performance Predictive Model.

To test the ability of the model to predict classifier accuracy a subset of the full 38 datasets are selected as the training set, with the remaining datasets left as hold-outs to evaluate the generality of the developed model. There are a large number of possible subsets for selecting 19 of the 38 datasets:

$$\binom{38}{19} = \frac{38!}{(38-19)!19!} = \frac{38!}{19!19!} = \frac{38 \cdot 37 \cdots 21 \cdot 20}{19!} = 35345263800 \tag{3.20}$$

Based on the impossibility of testing all combinations of the sample space, and the desire to try to cover a reasonable portion of the sample space, 400 prediction trials are executed. In each trial a linear model is generated in Matlab using the six difference measures as predictor variables and one of the combinations of classifier and performance measure as the response variable. Half of the 38 datasets are randomly selected as training datasets, and half as hold outs. The training datasets are then used to generate the linear model, with the hold-out datasets used to test the predictive ability of each linear model.

Once model generation is complete the distinct pairs of hold-out datasets in each trial have the following associated values for every initial classification repetition, classifier, and performance measure.

- Actual change in classifier performance
- Predicted change in classifier performance
- Difference between actual and predicted change in classifier performance (residuals)
- Model used

The ability of the selected difference measures to predict the change in classifier performance is evaluated by comparing the distribution of the residuals to the distribution of the actual classifier performance. For example, if a uniform distribution is assumed for both residuals and actual values, and if the range of the residuals covers 95% of the range of the actual classifier performance, then the predictor is giving very little information with respect to the actual value.

IV. Results

4.1 Correlation Within Datasets

The first set of experiments examine the possible correlation of the difference measured between a baseline dataset (d_i) and all other datasets, and the corresponding change in classifier performance. The hypotheses of this set of experiments follows.

Alternate Hypothesis 1. *The difference between a dataset and all other datasets will have a non-zero correlation with the corresponding changes in classifier performance.*

Null Hypothesis 1. *The difference between a dataset and all other datasets will have no correlation with the corresponding changes in classifier performance.*

The correlations were computed for each type of classifier, performance measure, and difference measure. All calculated differences between datasets and change in performance measures may be found in Appendix B and Appendix C, respectively. For each unique combination of those three factors, there were 2 sets of 38 correlations and two sets of 38 p-values. For example, the results for correlating δ_S (difference between sum of feature entropies) with Support Vector Machine (SVM) accuracy (A) are in Table 4.1.

Each correlation coefficient in Table 4.1 is calculated using one dataset (the *baseline* dataset) as the dataset which is being tested for any statistically significant non-zero correlation to the SVM accuracy. The correlation calculations, therefore, are based on the calculations of dataset difference and change in classifier performance from the dataset (d_i) being tested for utility to another dataset (d_j). For example, the value for Spearman's ρ for dataset two in Table 4.1 represents the rank correlation between two vectors of data. In the first vector (Equation (4.1)), each element is the difference measured between dataset two and the other datasets:

$$\vec{d}_2 = \langle \delta(H_2, H_1), \delta(H_2, H_3), \cdots, \delta(H_2, H_{37}), \delta(H_2, H_{38}) \rangle \tag{4.1}$$

Table 4.1: Correlations and p-values of δ_S and SVM accuracy for each baseline dataset.

Dataset	Spearman's ρ	p-value	Pearson's r	p-value	Dataset	Spearman's ρ	p-value	Pearson's r	p-value
1	0.3370	0.0414	0.3311	0.0453	20	0.3511	0.0331	0.3595	0.0289
2	0.3115	0.0606	0.2662	0.1112	21	0.4160	0.0104	0.3302	0.0460
3	0.2023	0.2298	0.1844	0.2746	22	0.2240	0.1827	0.1842	0.2752
4	0.2591	0.1215	0.2413	0.1502	23	0.2570	0.1247	0.2701	0.1060
5	0.1792	0.2886	0.2365	0.1587	24	0.4095	0.0118	0.3776	0.0212
6	0.2804	0.0927	0.2806	0.0925	25	0.2897	0.0820	0.3168	0.0561
7	0.2015	0.2318	0.2108	0.2103	26	0.3957	0.0153	0.3723	0.0233
8	0.4350	0.0071	0.3900	0.0170	27	0.4864	0.0023	0.4780	0.0028
9	0.2636	0.1149	0.2542	0.1289	28	0.2936	0.0778	0.2943	0.0770
10	0.3150	0.0576	0.3225	0.0516	29	0.3520	0.0327	0.4139	0.0109
11	0.3386	0.0404	0.3548	0.0312	30	0.3338	0.0435	0.3392	0.0400
12	0.3364	0.0418	0.2756	0.0987	31	0.2515	0.1333	0.2839	0.0886
13	0.1712	0.3109	0.2007	0.2337	32	0.1526	0.3673	0.1955	0.2463
14	0.3019	0.0694	0.2381	0.1558	33	0.1901	0.2598	0.1572	0.3527
15	0.3771	0.0214	0.2696	0.1065	34	0.3420	0.0383	0.3513	0.0330
16	0.3409	0.0389	0.3590	0.0291	35	0.3404	0.0392	0.3126	0.0596
17	0.2587	0.1220	0.2928	0.0787	36	0.3170	0.0559	0.2715	0.1041
18	0.3320	0.0447	0.3034	0.0680	37	0.3326	0.0443	0.3453	0.0363
19	0.3100	0.0619	0.3444	0.0369	38	0.3575	0.0298	0.3320	0.0447

Note that the difference between the *baseline* dataset and itself is omitted, as the difference will always be zero. In the second vector (Equation (4.2)), each element is the change in SVM classification accuracy from dataset dataset one to another dataset, calculated by subtracting the accuracy of SVM on other datasets from the accuracy on dataset two:

$$\vec{A}_2 = \langle A_2 - A_1, A_2 - A_3, \cdots, A_2 - A_{37}, A_2 - A_{38} \rangle \tag{4.2}$$

As the goal is to find a distance or difference measure which correlates to the changes in classifier performance, there is one indicator which, when combined with a p-value indicating statistically significant correlation, will demonstrate a consistent correlation between changes in difference and accuracy: the sign of the correlation coefficient. A

positive or negative correlation for all 38 datasets indicates the changes measured by the difference measure are related to the accuracy of the classifier regardless of the initial dataset. As a secondary concern, large magnitude correlations are a desired outcome: a consistently small correlation may be interesting, but larger magnitudes will be more useful for purposes of developing a predictive model.

Two conditions are necessary to reject the null hypothesis. First, the correlation must be statistically significant, with a p-value of less than 0.05. Second, the correlations for a given difference measure must be significantly different from zero. For the second to be meaningful, the first must be met: the null hypothesis cannot be rejected based on all correlations having the same sign if the results are not statistically significant. To determine if any of the results were sufficient to reject the null hypothesis it is necessary to calculate the number of results where all p-values are less than 0.05. There were no combinations of classifier, performance measure, and difference measure that met the first condition. The results with the most number of statistically significant results had only 33 of the 38 p-values (87%) less than 0.05. If there were many results which were significant for more than 85% of the datasets then they might merit further examination, however the distribution of values is such that most of the correlations were not significant. As shown in Figure 4.1, the distribution of significant p-value counts is such that most of the correlation results are not statistically significant.

Based on the results of the first set of experiments, there is insufficient evidence to reject Null Hypothesis 1 for any of the combinations of classifier, performance measure, and difference measure. There is insufficient evidence for the difference between a dataset and all other datasets having a non-zero correlation with the corresponding change in classifier performance.

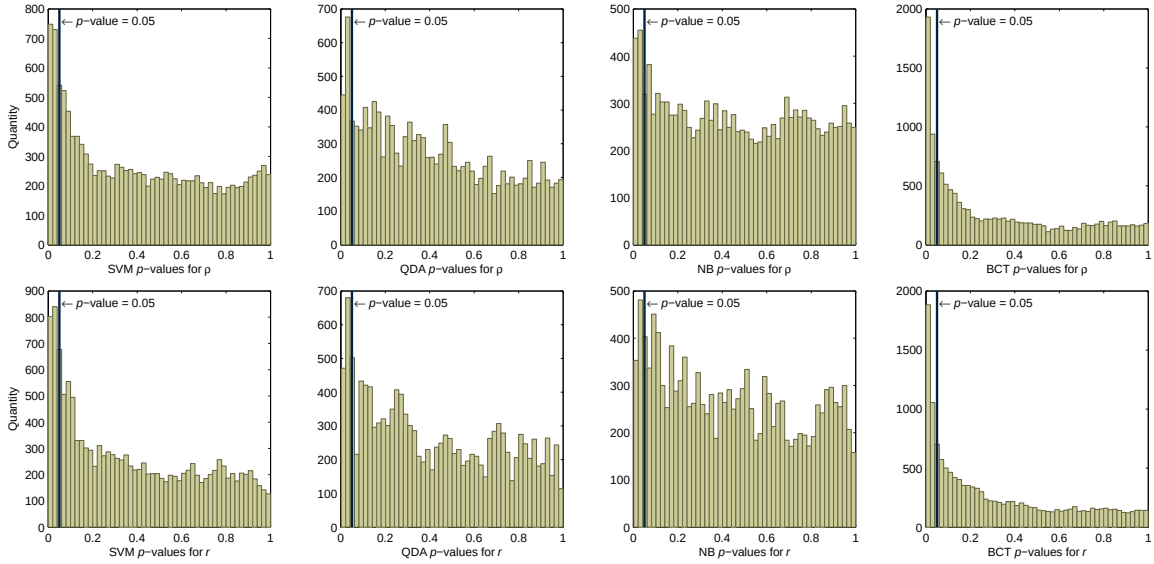

Figure 4.1: Histogram of Spearman's ρ (upper) and Pearson's r (lower) p-values. Histograms are per classifier, but combine all 20 repetitions and 18 possible pairings of difference and performance measures for each.

4.2 Correlation Across Datasets

Examination of the correlation of the differences between a single dataset and all other datasets and the corresponding change in classifier performance failed to produce sufficient evidence to reject the null hypothesis. However, the data can be examined in another way by combining the differences and classifier performance for all 703 unique combinations of datasets, and analyzing the correlation between each set of 703 differences and 703 changes in performance measure. The hypotheses of this set of experiments follows.

Alternate Hypothesis 2. *The difference between the unique pair-wise combinations of all 38 datasets will have a non-zero correlation with the corresponding changes in classifier performance.*

Null Hypothesis 2. *The difference between the unique pair-wise combinations of all 38 datasets will have no correlation with the corresponding changes in classifier performance.*

The correlations were computed for each type of classifier, performance measure, and difference measure. For each unique combination of those three factors, there were 20 repetitions of 2 correlations and twenty repetitions of 2 p-values.

Again, two conditions are necessary to reject the null hypothesis. First, the correlation must be statistically significant, with a p-value of less than 0.05. Second, the correlations for a given difference measure must be significantly different from zero. For the second to be meaningful, the first must be met: the null hypothesis cannot be rejected based on all correlations having the same sign if the results are not statistically significant. To determine if any of the results were sufficient to reject the null hypothesis it is necessary to calculate the number of results where all p-values are less than 0.05. While the p-values demonstrate that not all correlations in the results are statistically significant, they also indicate areas where the majority of the correlations *are* significant, and therefore which combinations of classifier and performance measure can be examined. There were multiple combinations of difference measure, classifier, and performance measure that met the first condition. Those combinations that did are shown in Table 4.2.

For every experiment in Table 4.2 there is sufficient evidence to reject Null Hypothesis 2. Therefore, there is evidence that the correlation of the difference between the unique pairwise combinations of all 38 datasets and the corresponding change in classifier performance is non-zero. Given the number of combinations that showed a significant correlation to classifier performance, it may be possible to predict the performance of a classifier knowing only the difference between the datasets.

4.3 Visual Examination

The Spearman and Pearson correlation coefficients are valid checks for correlation, but a visual check of a data is often good for verifying the results. In this case, the scatter and Quantile-Quantile (QQ) plots can provide insight into the shape of the results, including

Table 4.2: Combinations of difference, classifier, and performance with significant Spearman and Pearson correlation in all repetitions. Normality is tested using the Chi-square goodness of fit test, and difference from zero is tested using the one sample t-test.

Difference	Classifier	Performance	Normal	$\neq 0$	$\mu(\rho)$	$\sigma(\rho)$	$\mu(r)$	$\sigma(r)$
δ_J	SVM	A	True	True	0.3772	0.0037	0.3710	0.0035
δ_J	SVM	R_{P_f}	True	True	-0.2436	0.0342	-0.2710	0.0277
δ_J	QDA	A	True	True	-0.2875	0.0041	-0.2861	0.0040
δ_J	QDA	R_{P_t}	True	True	0.2147	0.0049	0.2434	0.0023
δ_J	NB	A	True	True	-0.0939	0.0047	-0.1010	0.0038
δ_J	NB	R_{P_t}	True	True	0.1977	0.0093	0.2306	0.0014
δ_J	BCT	R_{P_t}	True	True	0.2615	0.0151	0.2992	0.0145
δ_J	BCT	R_{P_f}	True	True	-0.1184	0.0133	-0.2127	0.0099
δ_S	SVM	A	True	True	0.3495	0.0043	0.3485	0.0038
δ_S	SVM	R_{P_f}	True	True	-0.2378	0.0288	-0.2373	0.0221
δ_S	QDA	A	True	True	-0.2224	0.0029	-0.2215	0.0031
δ_S	QDA	R_{P_t}	True	True	0.2746	0.0048	0.2932	0.0014
δ_S	NB	R_{P_t}	True	True	0.2915	0.0075	0.2867	0.0016
δ_S	NB	R_{P_f}	True	True	-0.1023	0.0027	-0.1480	0.0026
δ_S	BCT	A	True	True	0.3206	0.0202	0.3399	0.0203
δ_S	BCT	R_{P_t}	True	True	0.4628	0.0126	0.4413	0.0129
δ_S	BCT	R_{P_f}	True	True	-0.2240	0.0125	-0.2411	0.0098
$\delta_{\lvert F\rvert}$	NB	A	True	True	-0.1442	0.0020	-0.1426	0.0020
$\delta_{\lvert F\rvert}$	BCT	A	True	True	-0.1534	0.0089	-0.1723	0.0067
$\delta_{\lvert J\rvert}$	BCT	R_{P_t}	True	True	-0.0932	0.0061	-0.0847	0.0059
$\delta_{\lvert S\rvert}$	BCT	A	True	True	-0.1149	0.0092	-0.1274	0.0069

some patterns which become evident. The most obvious example of the pattern which occurs in some of the plots is shown in Figure 4.2.

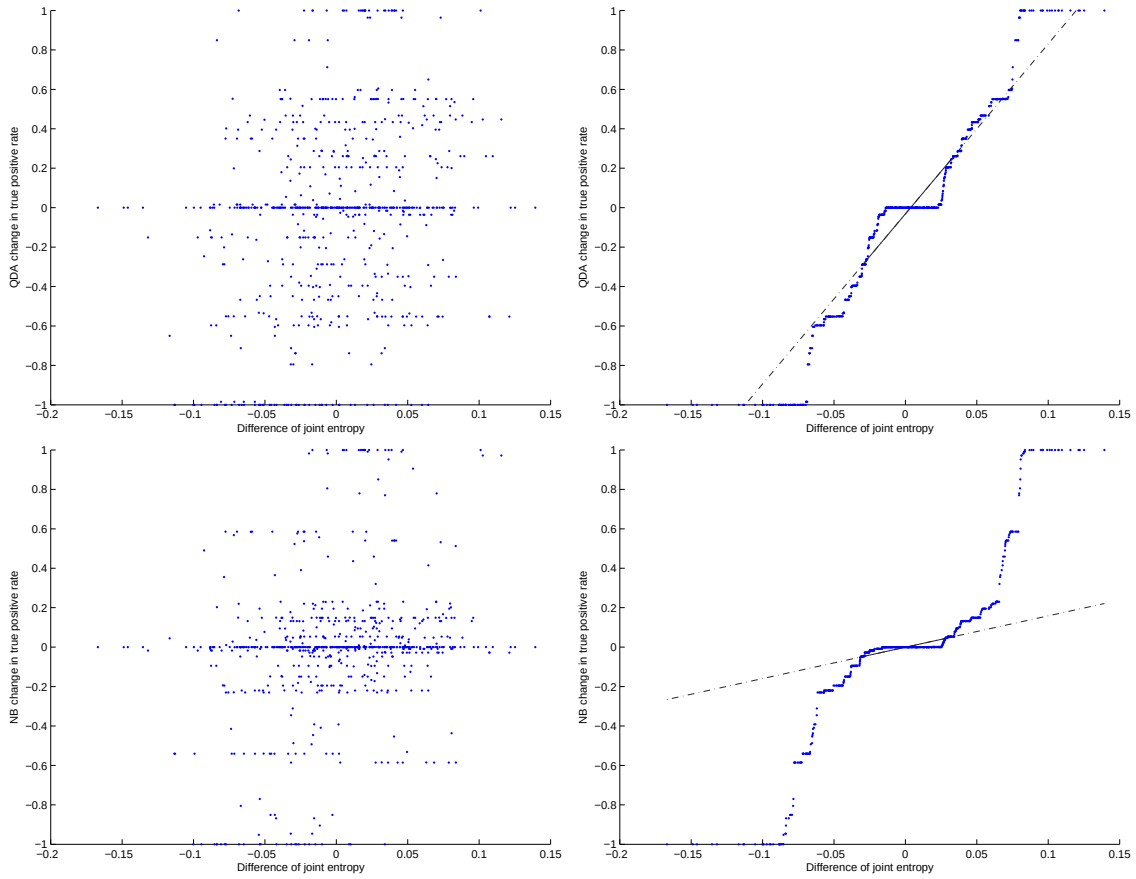

Figure 4.2: Striations echoed in both the scatter and QQ plots.

When present, the striation evident in Figure 4.2 is always in both the scatter and QQ plots. The displayed plots are the most obvious demonstrations of striation within the plots, but there are a total of seven of the 21 combinations which display the striations to a greater or lesser extent. The common striations in the scatter and QQ plots is further reason to be cautious in making claims based on these correlations. The other 14 combinations display a reasonable fit between the two distributions. There are minimal patterns within the scatter plots, and the QQ plots are fairly well matched: there are no perfect fits as shown by the three sets of plots showing a sampling from the remaining 14 combinations in Figure 4.3.

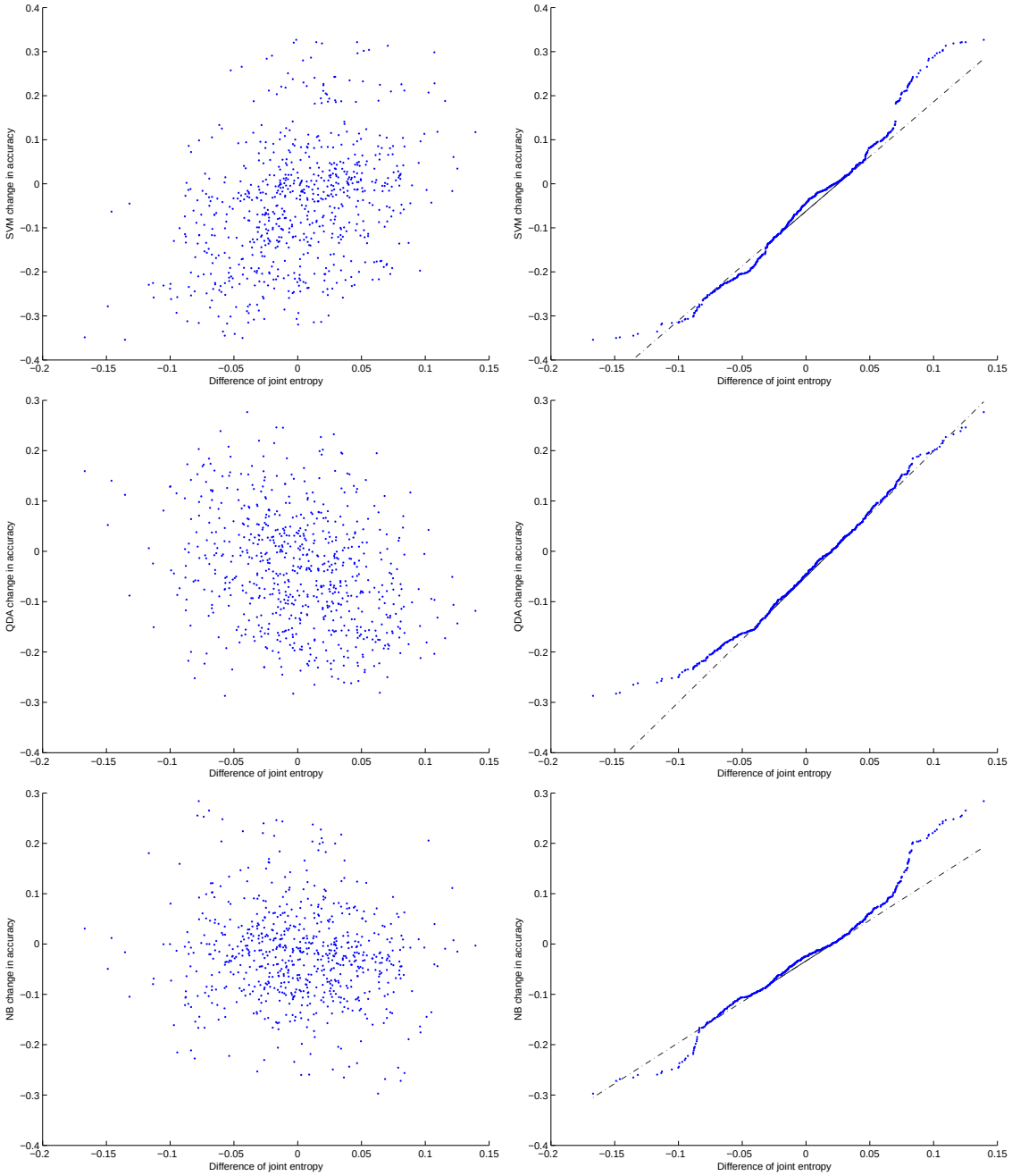

Figure 4.3: Sample of reasonable fits in both the scatter and QQ plots.

There is no type of difference (accuracy, true positive rate, and false positive rate) result that both correlates and is statistically significant for every classifier and dataset, although there are some that correlate for 36 or 37 of the datasets for a given classifier.

That is an indication that the difference measures chosen may not be the best measures for this particular use. The results are also an indication that more exploration may help in understanding why certain datasets result in correlations, and others do not.

The observation that a correlation exists between an entropy-based measure of difference and the change in classification results of all types of tested classifiers suggests that the prediction of classifier performance may be possible. Given a known dataset (d_i) and performance on that dataset, it may be possible to predict the classifier efficacy on a different dataset (d_j) based on the calculated difference between d_i to d_j. Given the different ways classifiers can be parameterized, and the methods which can be used to optimize results, it could be possible to provide estimates of classification accuracy, true positive rate, or false positive rate results.

4.4 Modeling Performance

The goal of this research is to produce a useful system to measure dataset difference that can be used to predict classifier performance change. Linear models are developed to predict classifier accuracy change based upon the difference measures calculated for each unique pair of datasets.

To test the ability of a linear model to predict classifier accuracy 400 prediction runs are executed. In each run a linear model is generated using the six difference measures as predictor variables and one of the combinations of classifier and performance measure as the response variable. Half of the 38 datasets are randomly selected as datasets to be used in training the linear model, and half as hold-outs. The training datasets are used to generate the linear model, and the hold-out datasets are used to test the predictive accuracy of the linear models.

Once complete the linear models are used to predict the change in classifier performance in the hold-out datasets. The actual performance change is subtracted from the predicted change to calculate the error for each set of prediction runs. Thereafter, every combination of

classifier and performance measure is associated with 400 sets of 171 actual, predicted, and error values. Individual runs may not be representative of the overall prediction capability of the model, therefore the probability distribution of values for all 400 runs of the actual, predicted, and errors are examined for each classifier.

Table 4.3: The minimum, 5^{th} percentile, median, 95^{th} percentile, and maximum prediction error of each classifier performance measure across all 400 prediction runs and 20 classification repetitions.

	Minimum	5^{th} percentile	Median	95^{th} percentile	Maximum
SVM accuracy	-0.47477	-0.25423	0.00550	0.22407	0.48619
QDA accuracy	-0.56089	-0.19656	0.00385	0.18675	0.38862
NB accuracy	-0.52048	-0.17528	0.00362	0.15882	0.35585
BCT accuracy	-0.10247	-0.04514	0.00110	0.04165	0.09603
SVM TPR	-0.62338	-0.29282	-0.01361	0.31020	0.75087
QDA TPR	-1.58474	-0.83083	-0.02096	0.90831	2.07642
NB TPR	-1.42219	-0.73996	0.00534	0.72415	1.36357
BCT TPR	-0.34932	-0.11881	-0.00078	0.13058	0.31741
SVM FPR	-0.54008	-0.29148	0.00465	0.28863	0.77028
QDA FPR	-2.14895	-1.10277	0.03038	1.19280	2.87723
NB FPR	-1.80074	-0.97441	0.01531	1.03935	2.08443
BCT FPR	-0.99866	-0.33455	0.00147	0.32241	0.91144

A large range among the errors can limit the utility of the predictions, so it is useful to determine which performance measure has the least error between predicted and actual performance, or *residual*, across all classifiers. To do so the minimum, 5^{th} percentile, median, 95^{th} percentile, and maximum value of each error distribution (across all 400 runs)

are calculated, and listed in Table 4.3. It is clear from the data in Table 4.3 that the models are most successful in predicting the accuracy performance measure. The True Positive Rate (TPR) and False Positive Rate (FPR) of some classifiers have 5^{th} percentile errors as large as -1.13 (Quadratic Discriminant Analysis (QDA) FPR). As an example of the effect large errors could have on prediction, an error of -1.13 would result in a predicted change in FPR of -0.23, while the actual change in FPR is 0.9. Given the large errors in TPR and FPR, accuracy is pursued as the performance measure most likely to be predicted with acceptable error.

Based on the distribution of residuals summarized in Table 4.3, accuracy is the most promising performance measure with respect to predicting classifier performance: for every classifier, the range of values between the 5^{th} and 95^{th} percentiles is less than the entire possible range. The residuals of each classifier are examined, and compared with the distribution of the actual classifier accuracy, to determine the predictive ability. Note that for the accuracy distributions in Figures 4.4, 4.5, 4.6 and 4.7 if the 95^{th} percentile line is above one, it is shown on the one. As an accuracy greater than one is not possible, there is no value in positioning the line on a value greater than one.

The first set of histograms examined are those in Figure 4.4 from the SVM classifier. The histogram of the actual classification values are shown to demonstrate the possible range of the predicted accuracy. Using the distribution of the SVM prediction residuals in Table 4.3, and assuming the predicted accuracy on the dataset being used is 0.85, the 90% confidence range of predicted accuracy would be from 0.60 to 1. The range of prediction residuals can cover the entire range of actual accuracy, resulting in no predictive ability.

The second set of histograms examined are those in Figure 4.5 from the QDA classifier. The histogram is again evaluated to determine whether the linear model provides any significant predictive ability. Based upon the range of the actual classification accuracies, the distribution of residuals, and assuming the predicted accuracy of 0.85, the 90% confidence

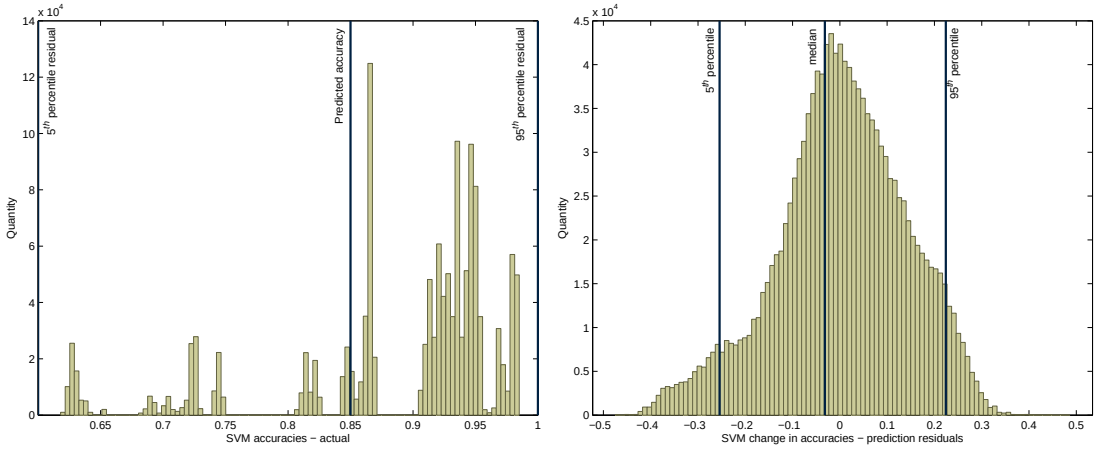

Figure 4.4: Histogram of 1,368,000 SVM classifier accuracies and error between predicted and actual accuracies, or *residuals*. The 1,368,000 is the result of the 171 unique pairings of 19 datasets, 20 classification repetitions, and 400 linear model predictive runs. The 90% confidence interval for the residuals is marked both on the residual plot, and on the accuracy plot assuming a predicted accuracy of 0.85.

range of predicted accuracy would be from 0.65 to 1. Again, the range of prediction residuals can cover the entire range of actual accuracy, resulting in no predictive ability.

The third set of histograms examined are those in Figure 4.6 from the Naïve Bayes (NB) classifier. The histogram is again evaluated to determine whether the linear model provides any significant predictive ability. Based upon the range of the actual classification accuracies, the distribution of residuals, and assuming the predicted accuracy of 0.85, the 90% confidence range of predicted accuracy would be from 0.67 to 1. Again, the range of prediction residuals can cover the entire range of actual accuracy, resulting in no predictive ability.

The fourth set of histograms examined are those in Figure 4.7 from the Binary Classification Tree (BCT) classifier. The histogram is again evaluated to determine whether

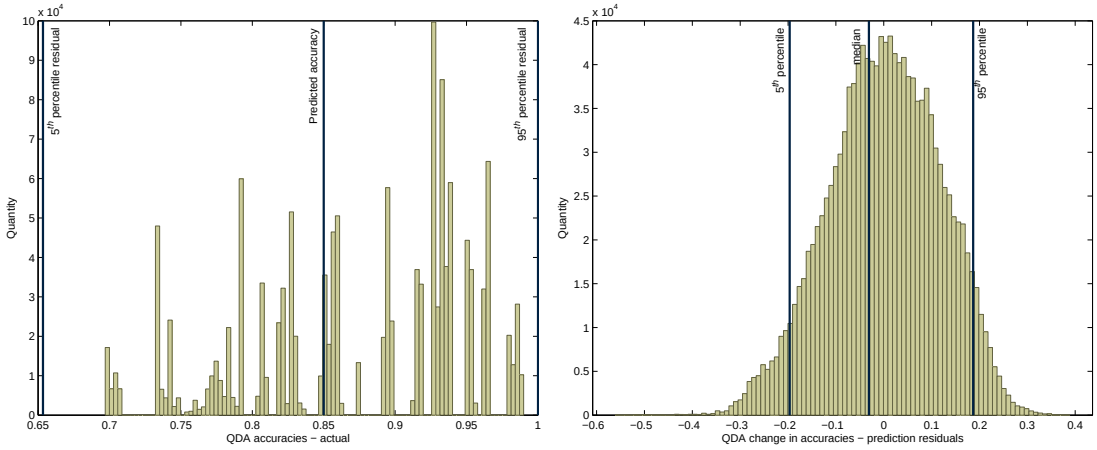

Figure 4.5: Histogram of 1,368,000 QDA classifier accuracies, and linear model prediction residuals. The 1,368,000 is the result of the 171 unique pairings of 19 datasets, 20 classification repetitions, and 400 linear model predictive runs. The 90% confidence interval for the residuals is marked both on the residual plot, and on the accuracy plot assuming a predicted accuracy of 0.85.

the linear model provides any significant predictive ability. Based upon the range of the actual classification accuracies, the distribution of residuals, and assuming the predicted accuracy of 0.96, the 90% confidence range of predicted accuracy would be from 0.91 to 1. In this case, the range of prediction does not quite cover the entire range of accuracies, failing to cover 0.0012 of the actual accuracy range, but there is still no predictive ability.

When comparing the results of predicting the change in accuracy of all four classifiers, shown in Figure 4.8, the linear model predictions for the change in SVM classifier accuracy have the widest range. This is also shown in Table 4.3, where the 5^{th} percentile is 0.05 below, and the 95^{th} percentile is 0.04 above, the nearest classifier. As the prediction results on SVM are the worst among all classifiers tested, any future efforts to reduce the size of prediction residuals should focus upon the prediction of change in SVM accuracy.

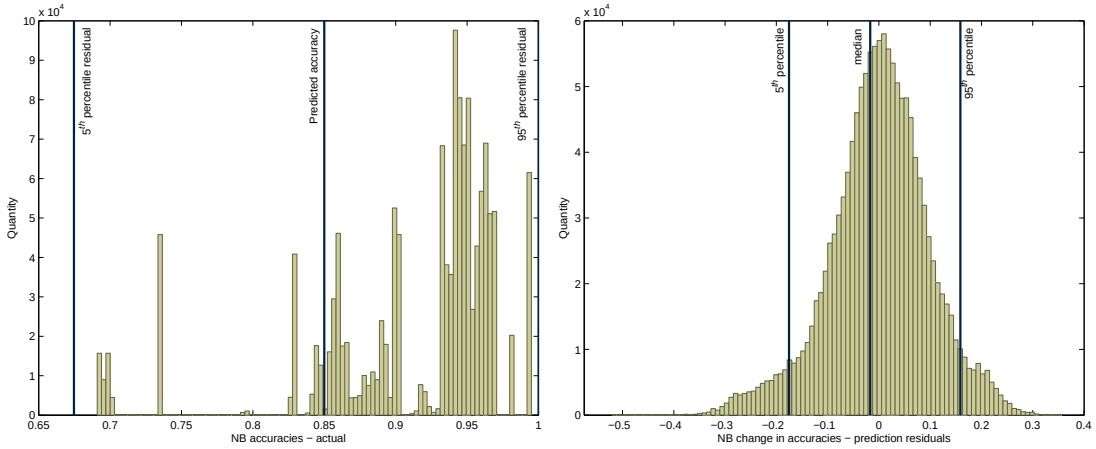

Figure 4.6: Histogram of $1,368,000$ NB classifier accuracies, and linear model prediction residuals. The $1,368,000$ is the result of the 171 unique pairings of 19 datasets, 20 classification repetitions, and 400 linear model predictive runs. The 90% confidence interval for the residuals is marked both on the residual plot, and on the accuracy plot assuming a predicted accuracy of 0.85.

For every classifier evaluated the distribution of the prediction residuals is such that predicted values could represent an actual value anywhere in the range of possible accuracies (see Table 4.4). Based on the evaluation of the developed linear models, there is no evidence that the developed models provide any significant predictive ability.

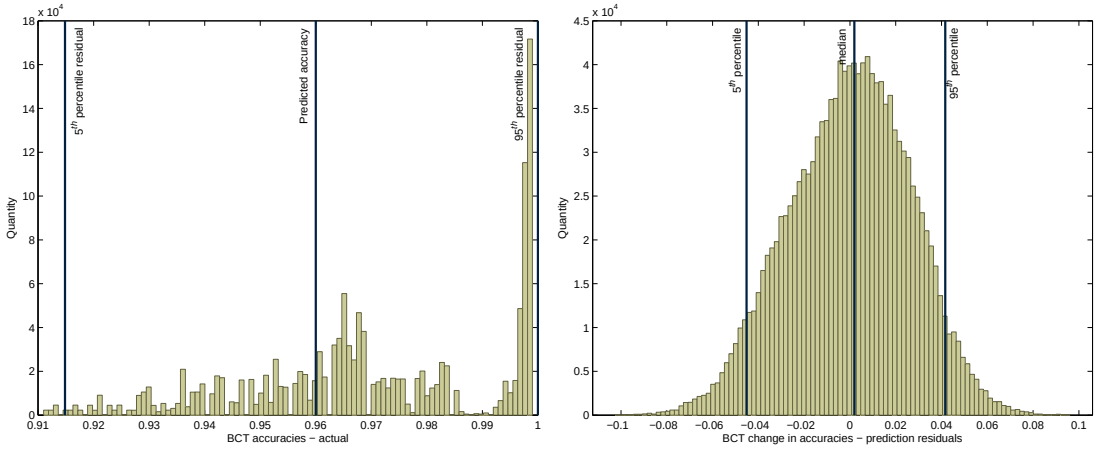

Figure 4.7: Histogram of $1,368,000$ BCT classifier accuracies, and linear model prediction residuals. The $1,368,000$ is the result of the 171 unique pairings of 19 datasets, 20 classification repetitions, and 400 linear model predictive runs. The 90% confidence interval for the residuals is marked both on the residual plot, and on the accuracy plot assuming a predicted accuracy of 0.96.

Table 4.4: Range of linear model prediction residuals and actual classifier accuracy.

	Range	
	Residuals	Actual Accuracy
SVM	0.4783	0.3670
QDA	0.3833	0.2930
NB	0.3341	0.3040
BCT	0.0868	0.0880

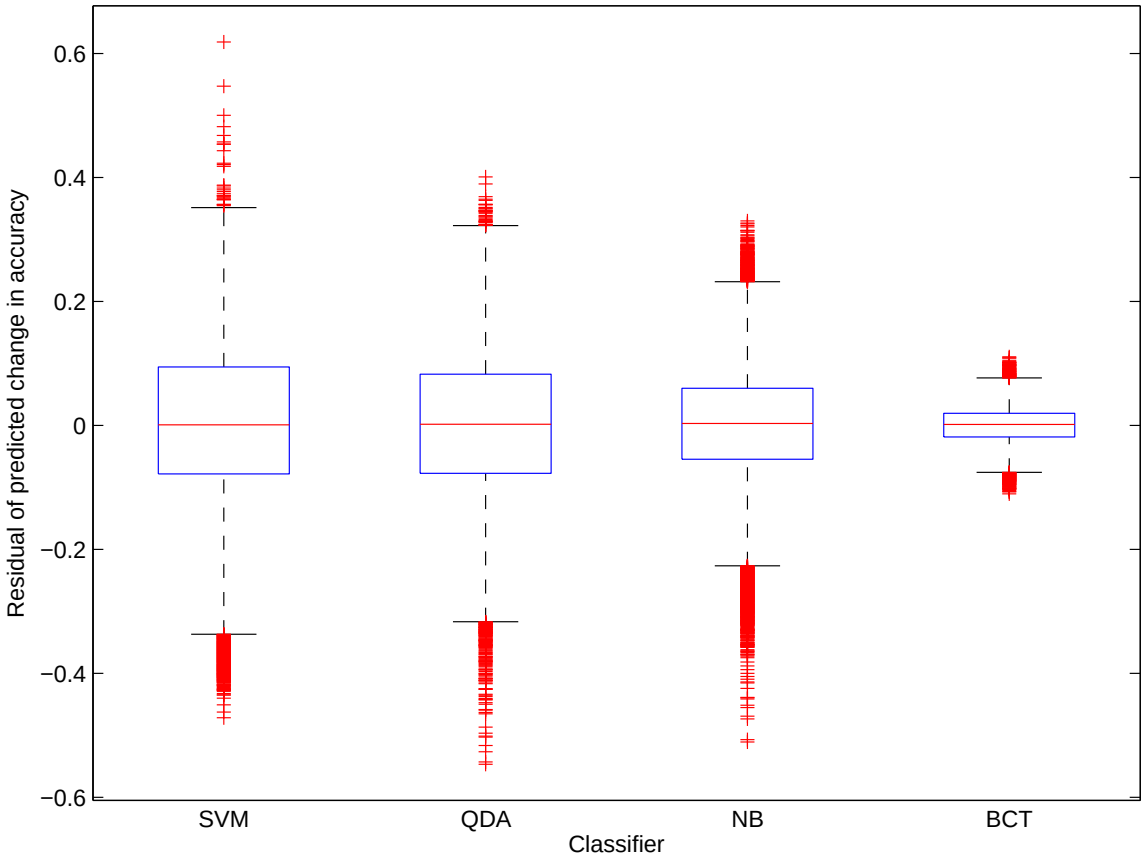

Figure 4.8: A comparison of the residuals of the accuracies predicted by the linear models for all four classifiers.

V. Conclusion

This chapter summarizes research into developing a system that predicts the change in classifier performance using linear models built from the differences measured between datasets, and the results presented in Chapter 4. Recommendations for future research are suggested by these results. The future work aims to improve the ability to model classifier performance as a function of differences between datasets.

5.1 Summary

The goal of this research was to develop a Classifier Accuracy Prediction System (CAPS) to predict the changes in classifier accuracy using the differences between datasets, and verify the prediction accuracy using the comparison of the distribution of the linear model prediction residuals and actual classifier accuracy over multiple prediction trials.

The first step in developing the CAPS was determining the type of difference measures which could be useful in predicting classifier accuracy, and determining the types of classifiers to be used. The difference measure types to be used were based upon the spatial and probabilistic measures available, although most were disparate types of entropy. Regarding selection of classifiers, a range were chosen to ensure broad coverage of the types of classifiers available to Network Intrusion Anomaly Detection (NIAD) developers and researchers. Correlations were performed to determine whether dataset-to-dataset correlations existed for all unique combinations of datasets, and it was determined they did not. Further examination of the differences between datasets and changes in classifier accuracy led to the calculation of correlation across datasets: correlating the differences of all unique combinations of datasets to the corresponding changes in accuracy. There was significant correlation among those values using two difference measures in particular: the difference between the sum of feature entropies, and the difference between joint entropy.

Despite consistent correlation, the linear models developed to predict classifier performance did not provide any evidence to support significant predictive ability.

While unable to progress to actual Network Intrusion (NI) datasets, this system of prediction provides a foundation upon which future work in the area of classifier performance prediction can be based.

5.2 Future Work

There are several areas in which the work presented here can be expanded. This section examines some of those areas for future research possibilities.

The results herein are based upon difference measures that are limited in scope, are not strongly correlated with the classifier performance changes, and do not fully explore the possible methods of measure difference. It will be useful for future researchers to methodically examine the use of different types of difference measures within the CAPS, to determine their utility for prediction of the change in classifier performance between datasets. It would also be useful to have a distance/similarity measure which had stronger consistent negative or positive correlations with the classification results, as that would allow better predications of the classification outcome based on distance from a common dataset.

The representation of difference as a scalar may not be as useful as a two, three, or n-dimensional representation of the distance or difference. Generating a vector within an entropy-space by using the entropy of each feature of a dataset may allow use of the distance while retaining information such as whether the distance is an increase or a decrease.

The number of features required for the change in accuracy to correlate with the difference between datasets is another metric which could provide some measure of how much information would be needed to represent a dataset effectively. Determining minimum number of features could give a lower-bound on the usefulness of any dataset for the purposes of predicting classifier performance, as any dataset without those features would be useless in predicting classifier performance.

The last recommendation for future work is related to the type of model used to predict accuracy. It is possible that the proper model to predict classifier accuracy is not linear in nature. Given that possibility, exploration of the different types of models available for predictive modeling would be of benefit. Identifying those models which were most useful may also give clues as to what other distance measure may prove useful as predictive terms.

These areas for future work are certainly not the only areas available. There are many others which have not been considered. Given the demonstrated correlation of the differences between datasets to the change in classifier performance on corresponding datasets, and the inability to predict the actual accuracy value using a linear model, continued research into this area using other models is indicated.

Appendix A: Papers Surveyed

- Shubair A. Abdulla, Sureswara Ramadass, Altyeb Altaher, and Amer Al Nassiri. "Setting a worm attack warning by using machine learning to classify netflow data." *International Journal of Computer Applications*, 36(2):49–56, December 2011.

- Saman M. Abdulla, Najla B. Al-Dabagh, and Omar Zakaria. "Identify features and parameters to devise an accurate intrusion detection system using artificial neural network." *World Academy of Science, Engineering and Technology*, 46(70):627–631, October 2010.

- Mohd Faizal Abdullah, Mohd Zaki Mas'ud, Shahrin Sahib, Robiah Yusof, Siti Rahayu Selamat, and Yaacob Asrul Hadi. "Time based intrusion detection on fast attack for network intrusion detection system." In *Second International Conference on Network Applications Protocols and Services*, pages 148–152, September 2010.

- Iftikhar Ahmad, Azween Abdullah, Abdullah Alghamdi, Khalid Alnfajan, and Muhammad Hussain. "Intrusion detection using feature subset selection based on MLP." *Scientific Research and Essays*, 6(34):6804–6810, December 2011.

- Iftikhar Ahmad, Azween Abdullah, Abdullah Alghamdi, and Muhammad Hussain. "Optimized intrusion detection mechanism using soft computing techniques." *Telecommunication Systems*, 47:1–9, July 2011.

- Iftikhar Ahmad, Azween B. Abdullah, and Abdullah S. Alghamdi. "Application of artificial neural network in detection of DOS attacks.". In *Proceedings of the Second International Conference on Security of Information and Networks*, pages 229–234, October 2009.

- Ghassan Ahmed Ali and Aman Jantan. "A new approach based on honeybee to improve intrusion detection system using neural network and bees algorithm." In *Software Engineering and Computer Systems*, volume 181 of *Communications in Computer and Information Science*, pages 777–792. Springer Berlin Heidelberg, 2011.

- Wafa' S. Al-Sharafat and Reyadh Naoum. "Significant of features selection for detecting network intrusions." In *International Conference for Internet Technology and Secured Transactions*, pages 1–6, November 2009.

- Hesham Altwaijry and Saeed Algarny. "Multi-layer bayesian based intrusion detection system." In *Proceedings of the World Congress on Engineering and Computer Science*, volume II of *WCECS 2011*, pages 918–922, 2011.

- Mário Antunes and Manuel Correia. "Tunable immune detectors for behaviour-based network intrusion detection." In *Artificial Immune Systems*, volume 6825 of *Lecture Notes in Computer Science*, pages 334–347. Springer Berlin / Heidelberg, 2011.

- Laleh Arshadi and Amir Hossein Jahangir. "Entropy based SYN flooding detection." In *Local Computer Networks (LCN), 2011 IEEE 36th Conference on*, pages 139–142, October 2011.

- Kusum Bharti and Sanyam Shukla and Shweta Jain. "Intrusion detection using unsupervised learning." *International Journal on Computer Science and Engineering*, 2(5):1865–1870, 2010.

- Samarjeet Borah, Saugat P. K. Chetry, and Pramod Kr Singh. "Hashed-K-means: A proposed intrusion detection algorithm." In *Computational Intelligence and Information Technology*, volume 250 of *Communications in Computer and Information Science*, pages 855–860. Springer Berlin Heidelberg, 2011.

- Imen Brahmi, Sadok Yahia, and Pascal Poncelet. "MAD-IDS: Novel intrusion detection system using mobile agents and data mining approaches." In *Intelligence and Security Informatics*, volume 6122 of *Lecture Notes in Computer Science*, pages 73–76. Springer Berlin / Heidelberg, 2010.

- Ye Changguo, Zhang Qin, Zhou Jingwei, Wei Nianzhong, Zhu Xiaorong, and Wang Tailei. "Improvement of association rules mining algorithm in wireless network intrusion detection." In *International Conference on Computational Intelligence and Natural Computing*, volume 2, pages 413–416, June 2009.

- Rung Ching Chen, Kai-Fan Cheng, and Chia-Fen Hsieh. "Using rough set and support vector machine for network intrusion detection." *International Journal of Network Security & Its Applications*, 1(1):1–13, April 2009.

- Xiaorong Cheng and Shanshan Wen. "A real-time hybrid intrusion detection system based on principle component analysis and self organizing maps." In *Sixth International Conference on Natural Computation*, volume 3, pages 1182–1185, August 2010.

- Te-Shun Chou, Kang K. Yen, and Jun Luo. "Network intrusion detection design using feature selection of soft computing paradigms." *International journal of computational intelligence*, 4(3):196–208, 2008.

- Vipin Das, Vijaya Pathak, Sattvik Sharma, Sreevathsan Ravichandran, MVVNS. Srikanth, and Kumar T. Gireesh "Network intrusion detection system based on machine learning algorithms." *International Journal of Computer Science & Information Technology*, 2(6):138–151, December 2010.

- David J. Day and Benjamin M. Burns. "A performance analysis of snort and suricata network intrusion detection and prevention engines." In *Fifth International Conference on Digital Society*, pages 187–192, February 2011.

- Nagaraju Devarakonda, Srinivasulu Pamidi, V. Valli Kumari, and A. Govardhan. "Outliers detection as network intrusion detection system using multi layered framework." In *Advances in Computer Science and Information Technology*, volume 131 of *Communications in Computer and Information Science*, pages 101–111. Springer Berlin Heidelberg, 2011.

- Entisar E. Eljadi and Zulaiha Ali Othman. "Anomaly detection for PTM's network traffic using association rule." In *3rd Conference on Data Mining and Optimization*, pages 63–69, June 2011.

- Vegard Engen, Jonathan Vincent, and Keith Phalp. "Exploring discrepancies in findings obtained with the KDD cup '99 data set." *Intelligent Data Analysis*, 15(2):251–276, April 2011.

- Robert Fanelli. "A hybrid model for immune inspired network intrusion detection." In *Artificial Immune Systems*, volume 5132 of *Lecture Notes in Computer Science*, pages 107–118. Springer Berlin / Heidelberg, 2008.

- Robert Fanelli. "Further experimentation with hybrid immune inspired network intrusion detection." In *Artificial Immune Systems*, volume 6209 of *Lecture Notes in Computer Science*, pages 264–275. Springer Berlin / Heidelberg, July 2010.

- Dewan Md. Farid and Mohammad Zahidur Rahman. "Attribute weighting with adaptive NBTree for reducing false positives in intrusion detection." *International Journal of Computer Science and Information Security*, 8(1):19–26, 2010. "abs/1005.0919.

- Dewan Md. Farid, Mohammad Zahidur Rahman, and Chowdhury Mofizur Rahman. Adaptive intrusion detection based on boosting and naïve bayesian classifier." "*International Journal of Computer Applications*, 24(3):12–19, June 2011. Published by Foundation of Computer Science.

- Dewan Md. Farid, Jérôme Darmont, Nouria Harbi, Huu Hoa Nguyen, and Mohammad Zahidur Rahman. "Adaptive network intrusion detection learning: Attribute selection and classification." In *Proceedings of the International Conference on Computer Systems Engineering*, Bangkok, Thailand, July 2009.

- Ed' Wilson Tavares Ferreira, Gilberto Arantes Carrijo, Ruy de Oliveira, and Nelcileno Araújo Virgílio de Souza. "Intrusion detection system with wavelet and neural artifical network approach for networks computers." *IEEE Latin America Transactions (Revista IEEE America Latina)*, 9(5):832–837, sept. 2011.

- Christiane Ferreira Lemos Lima, Francisco M. Assis, and Cleonilson Protásio de Souza. "A comparative study of use of Shannon, Rényi and Tsallis entropy for attribute selecting in network intrusion detection." In *2011 IEEE International Workshop on Measurements and Networking Proceedings*, pages 77–82, October 2011.

- Gianluigi Folino, Clara Pizzuti, and Giandomenico Spezzano. "An ensemble-based evolutionary framework for coping with distributed intrusion detection." *Genetic Programming and Evolvable Machines*, 11:131–146, June 2010.

- Jun Gao, Weiming Hu, Xiaoqin Zhang, and Xi Li. "Adaptive distributed intrusion detection using parametric model." In *IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies*, volume 1, pages 675–678, September 2009.

- Prasanta Gogoi, Bhogeswar Borah, and Dhruba K. Bhattacharyya. "Anomaly detection analysis of intrusion data using supervised & unsupervised approach." *Journal of Convergence Information Technology*, 5(1):95–110, February 2010.

- Prasanta Gogoi, Bhogeswar Borah, and Dhruba K. Bhattacharyya. "Network anomaly detection using unsupervised model." *International Journal of Computer Applications (Special Issue on Network Security and Cryptography)*, pages 19–30, 2011.

- Maoguo Gong, Jian Zhang, Jingjing Ma, and Licheng Jiao. "An efficient negative selection algorithm with further training for anomaly detection." *Knowledge-Based Systems*, 30:185–191, June 2012.

- Morteza Zi Hayat and Mahmoud Reza Hashemi. "An adaptive DCT based intrusion detection system." 2010.

- Di He. "Improving the computer network intrusion detection performance using the relevance vector machine with chebyshev chaotic map." In *IEEE International Symposium on Circuits and Systems*, pages 1584–1587, May 2011.

- Xin He and Sri Parameswaran. "MCAD: Multiple connection based anomaly detection." In *11th IEEE International Conference on Communication Systems*, pages 999–1004, November 2008.

- Weiming Hu, Wei Hu, and S. Maybank. "Adaboost-based algorithm for network intrusion detection." *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 38(2):577–583, April 2008.

- Marius Joldos and Ioan Lucian Muntean. "Distributed investigations of intrusion detection data on the grid." In *10th RoEduNet International Conference*, pages 1–4, June 2011.

- M. Sadiq Ali Khan and S. M. Aqil Burney. "Efficient FSM techniques for IDS to reduce the system attacks." *International Journal of Computer Applications*, 29(11):42–47, September 2011.

- Kok-Chin Khor, Choo-Yee Ting, and Somnuk Phon-Amnuaisuk. "A cascaded classifier approach for improving detection rates on rare attack categories in network intrusion detection." *Applied Intelligence*, 36(2):320–329, March 2012.

- Gang Kou, Yi Peng, Zhengxin Chen, and Yong Shi. "Multiple criteria mathematical programming for multi-class classification and application in network intrusion detection." *Information Sciences*, 179(4):371–381, 2009.

- Farah Barika Ktata, Nabil El Kadhi, and Khaled Ghédira. "Agent IDS based on misuse approach." *Journal of Software*, 4(6):495–507, August 2009.

- Farah Barika Ktata, Nabil El Kadhi, and Khaled Ghédira. "MA_IDS: Mobile agents for intrusion detection system." In *IEEE International Advance Computing Conference*, pages 900–910, March 2009.

- Yinhui Li, Jingbo Xia, Silan Zhang, Jiakai Yan, Xiaochuan Ai, and Kuobin Dai. "An efficient intrusion detection system based on support vector machines and gradually feature removal method." *Expert Systems with Applications*, 39(1):424–430, January 2012.

- Liang Fu Lu, Mao Lin Huang, M.A. Orgun, and Jia Wan Zhang. "An improved wavelet analysis method for detecting DDoS attacks." In *4th International Conference on Network and System Security*, pages 318–322, September 2010.

- Shingo Mabu, Ci Chen, Nannan Lu, K. Shimada, and K. Hirasawa. "An intrusion-detection model based on fuzzy class-association-rule mining using genetic network

programming." *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 41(1):130–139, January 2011.

- Rasha G. Mohammed and Awad M. Awadelkarim. "Design and implementation of a data mining-based network intrusion detection scheme." *Asian Journal of Information Technology*, 10(4):136–141, 2011.

- Antti Niemelä. "Traffic analysis for intrusion detection in telecommunications networks." Master's thesis, Tampere University of Technology, March 2011.

- Mohammad Reza Norouzian and Sobhan Merati. "Classifying attacks in a network intrusion detection system based on artificial neural networks." In *13th International Conference on Advanced Communication Technology*, pages 868–873, February 2011.

- Charlie Obimbo, Haochen Zhou, and Ryan Wilson. "Multiple SOFMs working cooperatively in a vote-based ranking system for network intrusion detection." *Procedia Computer Science (Special Issue on Complex adaptive sysytems)*, 6:219–224, 2011.

- Agustin Orfila, Juan Estevez-Tapiador, and Arturo Ribagorda. "Evolving high-speed, easy-to-understand network intrusion detection rules with genetic programming." In *Applications of Evolutionary Computing*, volume 5484 of *Lecture Notes in Computer Science*, pages 93–98. Springer Berlin / Heidelberg, 2009.

- Francesco Palmieri and Ugo Fiore. "Network anomaly detection through nonlinear analysis." *Computers & Security*, 29(7):737–755, October 2010.

- Mrutyunjaya Panda and Manas Ranjan Patra. "Ensemble of classifiers for detecting network intrusion." In *Proceedings of the International Conference on Advances in Computing, Communication and Control*, pages 510–515, January 2009.

- Sergio Pastrana, Agustin Orfila, and Arturo Ribagorda. “A functional framework to evade network IDS.” In *44th Hawaii International Conference on System Sciences*, pages 1–10, January 2011.

- Martin Rehak, Michal Pechoucek, Pavel Celeda, Jiri Novotny, and Pavel Minarik. “CAMNEP: agent-based network intrusion detection system.” In *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems: Industrial Track*, pages 133–136, 2008.

- Mostafa Salama, Heba Eid, Rabie Ramadan, Ashraf Darwish, and Aboul Hassanien. “Hybrid intelligent intrusion detection scheme.” In *Soft Computing in Industrial Applications*, volume 96 of *Advances in Intelligent and Soft Computing*, pages 293–303. Springer Berlin / Heidelberg, 2011.

- Hadi Sarvari and Mohammad Mehdi Keikha. “Improving the accuracy of intrusion detection systems by using the combination of machine learning approaches.” In *International Conference of Soft Computing and Pattern Recognition*, pages 334–337, December 2010.

- Naeem Seliya and Taghi M. Khoshgoftaar. “Active learning with neural networks for intrusion detection.” In *IEEE International Conference on Information Reuse and Integration*, pages 49–54, August 2010.

- Sevil Sen and John A. Clark. “Evolutionary computation techniques for intrusion detection in mobile ad hoc networks.” *Computer Networks*, 55(15):3441–3457, October 2011.

- Kamran Shafi and Hussein Abbass. “Evaluation of an adaptive genetic-based signature extraction system for network intrusion detection.” *Pattern Analysis & Applications*, pages 1–18, November 2011.

- Bharanidharan Shanmugam and Norbik Bashah Idris. "Improved intrusion detection system using fuzzy logic for detecting anamoly and misuse type of attacks." In *International Conference of Soft Computing and Pattern Recognition*, pages 212–217, December 2009.

- Mansour Sheikhan, Zahra Jadidi, and Maedeh Beheshti. "Effects of feature reduction on the performance of attack recognition by static and dynamic neural networks." *World Applied Sciences Journal*, 8(3):302–308, 2010.

- Mansour Sheikhan and Amir Ali Sha'bani. "Fast neural intrusion detection system based on hidden weight optimization algorithm and feature selection." *World Applied Sciences Journal (Special Issue of Computer & IT)*, 7:45–53, 2009.

- Mei-Ling Shyu and Varsha Sainani. "A multiagent-based intrusion detection system with the support of multi-class supervised classification." In *Data Mining and Multi-agent Integration*, pages 127–142. Springer US, 2009.

- Siva S. Sivatha Sindhu, S. Geetha, and A. Kannan. "Decision tree based light weight intrusion detection using a wrapper approach." *Expert Systems with Applications*, 39(1):129–141, 2012.

- Shailendra Singh and Sanjay Silakari. "An ensemble approach for feature selection of cyber attack dataset." *International Journal of Computer Science and Information Security*, 6(2):297–302, 2009. abs/0912.1014.

- Shailendra Singh and Sanjay Silakari. "Generalized discriminant analysis algorithm for feature reduction in cyber attack detection system." *International Journal of Computer Science and Information Security*, 6(1):173–180, October 2009. abs/0911.0787.

- Jungsuk Song, Hiroki Takakura, Yasuo Okabe, and Koji Nakao. "Toward a more practical unsupervised anomaly detection system." *Information Sciences*, 2011. (In press).

- Pamidi Srinivasulu, Jalleda Ranga Rao, and Inampudi Ramesh Babu. "Network intrusion detection using FP tree rules." *Journal Of Advanced Networking and Applications*, 1(1):30–39, 2009.

- Ming-Yang Su, Gwo-Jong Yu, and Chun-Yuen Lin. "A real-time network intrusion detection system for large-scale attacks based on an incremental mining approach." *Computers & Security*, 28(5):301–309, 2009.

- Zhiyuan Tan, Aruna Jamdagni, Xiangjian He, and Priyadarsi Nanda. "Network intrusion detection based on LDA for payload feature selection." In *IEEE GLOBECOM Workshops*, pages 1545–1549, December 2010.

- Rubana Tarannum and Megha Lamble. "Hybrid approach: Detection of intrusion in manet." *IJCA Proceedings on Innovative Conference on Embedded Systems, Mobile Communication and Computing*, ICEMC2(1):24–28, September 2011.

- Jingwen Tian and Meijuan Gao. "Network intrusion detection method based on high speed and precise genetic algorithm neural network." In *International Conference on Networks Security, Wireless Communications and Trusted Computing*, volume 2, pages 619–622, April 2009.

- Carmen Torrano-Giménez, Alejandro Pérez-Villegas, Gonzalo Álvarez Marañón, et al. "An anomaly-based approach for intrusion detection in web traffic." *Journal of Information Assurance and Security*, 5(4):446–454, 2010.

- Taner Tuncer and Yetkin Tatar. "Detection DoS attack on FPGA using fuzzy association rules." In *IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 1271–1276, November 2011.

- R. Vijayasarathy, Balaraman Ravindran, and S.V. Raghavan. "A system approach to network modeling for DDoS detection using a naïve bayesian classifier." In *Third International Conference on Communication Systems and Networks*, pages 1–10, January 2011.

- J. Visumathi and K. L. Shunmuganathan. "A computational intelligence for evaluation of intrusion detection system." *Indian Journal of Science and Technology*, 4(1):40–45, January 2011.

- Juan Wang, Qiren Yang, and Dasen Ren. "An intrusion detection algorithm based on decision tree technology." In *Asia-Pacific Conference on Information Processing*, volume 2, pages 333–335, July 2009.

- Jun Wang, Taihang Li, and Rongrong Ren. "A real time IDSs based on artificial bee colony-support vector machine algorithm." In *Third International Workshop on Advanced Computational Intelligence*, pages 91–96, August 2010.

- Yong Wang, Dawu Gu, Mi Wen, Jianping Xu, and Haming Li. "Denial of service detection with hybrid fuzzy set based feed forward neural network." In *Advances in Neural Networks - ISNN 2010*, volume 6064 of *Lecture Notes in Computer Science*, pages 576–585. Springer Berlin / Heidelberg, June 2010.

- Jing Xiao-Pei and Wang Hou-Xiang. "A new immunity intrusion detection model based on genetic algorithm and vaccine mechanism." *International Journal of Computer Network and Information Security*, 2(2):33–39, December 2010.

- Qinzhen Xu, Zhimao Bai, and Luxi Yang. "An improved perceptron tree learning model based intrusion detection approach." In *International Conference on Artificial Intelligence and Computational Intelligence*, volume 4, pages 307–311, November 2009.

- Yun Yang and Jia Mi. "Design and implementation of distributed intrusion detection system based on honeypot." In *2nd International Conference on Computer Engineering and Technology*, volume 6, pages V6–260–V6–263, April 2010.

- Xia Ye, Junshan Li, and Yanling Li. "An anomaly detection system based on hide markov model for MANET." In *Sixth International Conference on Wireless Communications Networking and Mobile Computing*, pages 1–4, September 2010.

- Yang Yi, Jiansheng Wu, and Wei Xu. "Incremental SVM based on reserved set for network intrusion detection." *Expert Systems with Applications*, 38(6):7698–7707, June 2011.

- Wang Yu, Cheng Xiaohui, and Wang Sheng. "Anomaly network detection model based on mobile agent." In *Third International Conference on Measuring Technology and Mechatronics Automation*, volume 1, pages 504–507, January 2011.

- Anazida Zainal, Mohd Aizaini Maarof, Siti Mariyam Shamsuddin. "Ensemble classifiers for network intrusion detection system." *Journal of Information Assurance and Security*, 4:217–225, 2009.

- Safaa Zaman and Fakhri Karray. "Fuzzy ESVDF approach for intrusion detection systems." In *International Conference on Advanced Information Networking and Applications*, pages 539–545, May 2009.

- Mahdi Zamani, Mahnush Movahedi, Mohammad Ebadzadeh, and Hossein Pedram. "A DDoS-aware IDS model based on danger theory and mobile agents." In *International*

Conference on Computational Intelligence and Security, volume 1, pages 516–520, December 2009.

- Zargar, Gholam Reza and Kabiri, Peyman "Identification of effective network features to detect smurf attacks." In *IEEE Student Conference on Research and Development*, pages 49–52, November 2009.

- Bin Zeng, Lu Yao, and ZhiChen Chen. "A network intrusion detection system with the snooping agents." In *International Conference on Computer Application and System Modeling*, volume 3, pages V3–232–V3–236, October 2010.

- Biying Zhang. "A heuristic genetic neural network for intrusion detection." In *International Conference on Internet Computing Information Services*, pages 510–513, September 2011.

- Cheng Zhang, Jing Zhang, Sunjun Liu, and Yintian Liu. "Network intrusion active defense model based on artificial immune system." In *Fourth International Conference on Natural Computation*, volume 1, pages 97–100, October 2008.

- Hongying Zheng, Meiju Hou, and Yu Wang. "An efficient hybrid clustering-PSO algorithm for anomaly intrusion detection." *Journal of Software*, 6(12):2350–2360, December 2011.

- Hu Zhengbing, Li Zhitang, and Wu Junqi. "A novel network intrusion detection system (NIDS) based on signatures search of data mining." In *First International Workshop on Knowledge Discovery and Data Mining*, pages 10–16, January 2008.

- Ma Zhenying. "Reason for hierarchical self organized map-based intrusion detection system incapable of increasing detection rate." In *International Symposium on Information Engineering and Electronic Commerce*, pages 150–154, May 2009.

- Fangzhou Zhu, Jun Long, Wentao Zhao, and Zhiping Cai. “A misleading attack against semi-supervised learning for intrusion detection.” In VicenÃğ Torra, Yasuo Narukawa, and Marc Daumas, editors, *Modeling Decisions for Artificial Intelligence*, volume 6408 of *Lecture Notes in Computer Science*, pages 287–298. Springer Berlin / Heidelberg, 2010.

- Zhenyun Zhuang, Ying Li, and Zesheng Chen. “PAIDS: A proximity-assisted intrusion detection system for unidentified worms.” In *33rd Annual IEEE International Computer Software and Applications Conference*, volume 1, pages 392–399, July 2009.

Appendix B: Differences Between Datasets

Table B.1: Differences between δ_J of each dataset.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	0.0000	-0.0702	-0.0480	-0.0306	0.0321	-0.0656	-0.0752	-0.0335	-0.0486	-0.0418	-0.0963	-0.1010	-0.0464	-0.0552	0.0144	-0.0353	-0.0529	-0.0965	-0.0674	-0.0499	-0.0149	-0.0796	-0.0210	-0.0374	-0.0495	-0.0387	0.0023	-0.0232	0.0568	0.0940	-0.0344	-0.0869	-0.0210	-0.0419	-0.0740	-0.0725	-0.0957	-0.0958
2	0.0702	0.0000	0.0884	0.0253	0.0771	0.0217	-0.0083	0.0856	0.0091	0.0293	-0.0133	0.0060	0.0151	0.0504	0.1319	0.0522	0.0257	0.0319	0.0196	0.0147	0.0872	-0.0066	0.0717	0.0343	0.0345	0.0191	0.0970	0.0750	0.0803	0.0620	0.0160	-0.0032	0.0877	0.0352	0.0231	0.0018	0.0151	-0.0128
3	0.0480	-0.0884	0.0000	-0.0159	0.0236	-0.0665	-0.0813	-0.0236	-0.0132	-0.0731	-0.0469	-0.0456	-0.0161	-0.0129	0.0304	-0.0186	-0.0381	-0.0446	-0.0804	-0.0797	0.0134	-0.0307	0.0273	-0.0203	-0.0518	-0.0220	0.0155	-0.0472	0.0533	-0.0585	-0.0610	-0.0709	-0.0071	0.0077	-0.0616	-0.0279	-0.0509	-0.0610
4	0.0306	-0.0253	0.0159	0.0000	0.0728	0.0096	0.0076	0.0226	0.0439	-0.0407	-0.0145	-0.0200	0.0615	-0.0150	0.0404	0.0100	0.0093	0.0422	-0.0097	-0.0469	0.0039	-0.0436	0.0449	-0.0147	0.0011	-0.0045	0.0249	0.0857	0.0880	0.0410	0.0038	-0.0083	0.0220	0.0091	-0.0298	0.0050	-0.0032	-0.0366
5	-0.0321	-0.0771	-0.0236	-0.0726	0.0000	-0.0601	-0.1073	-0.0633	-0.0286	-0.0836	-0.1071	-0.1155	-0.0321	-0.0026	0.0189	-0.0747	-0.0465	-0.0303	-0.0707	-0.0802	-0.0386	-0.0449	0.0120	-0.1211	-0.0559	-0.1025	-0.0464	-0.0512	0.0167	-0.0344	-0.0744	-0.0643	-0.0016	-0.0505	-0.0494	-0.0263	-0.0716	-0.0583
6	0.0636	-0.0217	0.0665	-0.0096	0.0601	0.0000	-0.0294	0.0164	0.0525	-0.0202	-0.0706	-0.0229	0.0595	-0.0169	0.0855	0.0019	0.0148	0.0138	-0.0196	-0.0291	0.0284	0.0254	0.0142	-0.0131	-0.0390	-0.0071	0.0105	0.0382	0.0420	0.0197	-0.0077	-0.0278	0.0693	-0.0070	0.0002	0.0110	0.0151	-0.0259
7	0.0752	0.0083	0.0813	-0.0076	0.1073	0.0294	0.0000	0.0684	0.0567	-0.0391	-0.0063	-0.0157	0.0650	0.0359	0.0713	0.0158	0.0279	0.0522	-0.0012	0.0075	0.0312	0.0155	0.0599	-0.0020	0.0070	-0.0053	0.0615	0.0673	0.1489	0.0375	-0.0258	-0.0046	0.1052	0.0184	0.0189	0.0233	0.0071	0.0003
8	0.0335	-0.0836	0.0216	-0.0226	0.0633	-0.0164	-0.0684	0.0000	-0.0022	-0.0276	-0.0644	-0.0702	0.0103	-0.0519	0.0282	-0.0560	-0.0360	-0.0919	-0.0419	-0.0158	0.0117	-0.0268	0.0430	-0.0216	-0.0261	-0.0162	0.0541	0.0571	0.0663	0.0514	-0.0524	-0.0317	0.0164	-0.0245	-0.0599	-0.0222	-0.0131	-0.0283
9	0.0486	-0.0991	0.0132	-0.0439	0.0286	-0.0525	-0.0567	0.0022	0.0000	-0.0406	0.0088	-0.0289	-0.0162	-0.0283	0.0439	-0.0079	-0.0020	-0.0063	-0.0445	-0.0044	-0.0106	-0.0619	0.0268	0.0068	-0.0089	-0.0197	0.0273	0.0483	0.0930	0.0398	-0.0648	-0.0656	0.0031	-0.0226	-0.0424	-0.0199	-0.0354	-0.0271
10	0.0418	-0.0293	0.0731	0.0407	0.0836	0.0202	0.0391	0.0276	0.0406	0.0000	0.0025	-0.0263	0.0570	0.0162	0.0433	0.0317	-0.0019	0.0155	0.0337	0.0305	0.0732	0.0173	0.1167	0.0647	0.0466	0.0059	0.1134	0.0714	0.0993	0.1129	0.0071	0.0285	0.0569	0.0301	0.0224	0.0109	0.0456	-0.0015
11	0.0963	0.0133	0.0469	0.0145	0.1071	0.0706	0.0963	0.0644	-0.0088	-0.0926	0.0000	0.0067	0.0000	0.0419	0.0116	0.0773	0.0670	0.0279	0.0431	0.0100	0.0536	0.0316	0.0776	0.0387	0.0273	0.0168	0.0865	0.0586	0.1669	0.0572	-0.0099	0.0337	0.0527	0.0344	0.0305	0.0190	0.0421	0.0221	0.0461
12	0.1010	-0.0060	0.0456	0.0200	0.1135	0.0229	0.0157	0.0702	0.0289	0.0263	-0.0067	0.0000	0.0419	0.0116	0.0773	0.0670	0.0279	0.0431	0.0100	0.0536	0.0543	0.0737	0.0681	0.0348	0.0184	0.0837	0.0948	0.1001	0.0853	0.0514	-0.0054	0.0288	0.0530	0.0319	0.0190	0.0421	0.0221	0.0461
13	0.0464	-0.0151	0.0161	-0.0615	0.0321	-0.0595	-0.0650	-0.0103	0.0162	-0.0570	-0.0219	-0.0419	0.0000	-0.0590	0.0449	-0.0300	-0.0146	-0.0094	-0.0682	-0.0282	0.0138	-0.0883	0.0223	-0.0049	-0.0247	-0.0466	0.0232	0.0698	0.0673	0.0573	-0.0363	-0.1047	-0.0349	-0.0334	0.0464	-0.0245	-0.0387	0.0163	-0.0183	-0.0445
14	0.0552	-0.0504	0.0129	0.0150	0.0026	0.0169	-0.0359	0.0519	0.0283	-0.0162	-0.0396	-0.0116	0.0390	0.0000	0.0841	0.0138	-0.0639	0.0250	-0.0476	-0.0083	0.0408	0.0039	0.0926	0.0195	0.0129	-0.0537	0.0045	0.0035	0.0877	0.0262	-0.0526	-0.0268	0.0298	0.0044	-0.0074	0.0560	-0.0363	0.0142
15	-0.0144	-0.1319	-0.0304	-0.0404	-0.0180	-0.0855	-0.0713	-0.0282	-0.0439	-0.0433	-0.0997	-0.0773	-0.0449	-0.0841	0.0000	-0.0791	-0.0813	-0.0720	-0.0831	-0.0698	-0.0153	-0.0681	-0.0115	-0.0405	-0.0511	-0.0567	-0.0514	-0.0205	0.0127	-0.0491	-0.0796	-0.0896	-0.0294	-0.0290	-0.0749	-0.0734	-0.0657	-0.0806
16	0.0335	-0.0322	0.0186	-0.0100	0.0747	-0.0019	-0.0158	0.0340	0.0079	-0.0317	-0.0538	-0.0670	0.0300	-0.0138	0.0791	0.0000	0.0105	0.0183	-0.0203	-0.0269	0.0034	0.0384	0.0248	0.0130	0.0095	0.0063	0.0304	0.0479	0.0784	0.0373	-0.0039	-0.0027	0.0794	-0.0197	0.0020	0.0024	0.0066	-0.0077
17	0.0529	-0.0257	0.0381	-0.0093	0.0465	-0.0148	-0.0279	0.0360	0.0020	0.0019	-0.0318	-0.0279	0.0146	0.0659	0.0813	-0.0105	0.0000	0.0307	-0.0159	0.0092	0.0250	0.0085	0.0590	-0.0131	-0.0041	-0.0188	0.0287	0.0817	0.0885	0.0653	-0.0306	-0.0266	0.0098	0.0009	-0.0115	-0.0019	-0.0162	0.0057
18	0.0695	-0.0319	0.0446	-0.0422	0.0303	-0.0138	-0.0522	0.0019	0.0063	-0.0135	-0.0687	-0.0431	0.0094	-0.0250	0.0720	-0.0183	-0.0307	0.0000	-0.0662	-0.0301	0.0141	0.0048	-0.0118	-0.0219	-0.0004	-0.0293	0.0379	0.0360	0.0247	0.0194	-0.0479	-0.0334	0.0464	-0.0245	-0.0087	0.0163	-0.0183	-0.0445
19	0.0674	-0.0196	0.0804	0.0097	0.0307	0.0196	0.0012	0.0419	0.0445	-0.0337	-0.0104	-0.0100	0.0682	0.0476	0.0831	0.0203	0.0159	0.0662	0.0000	0.0027	0.0000	0.0773	0.0268	0.0786	-0.0077	0.0243	0.0471	0.0309	0.0756	0.0659	-0.0285	-0.0358	0.0718	0.0134	-0.0121	-0.0018	0.0395	0.0114
20	0.0699	-0.0147	0.0797	0.0469	0.0082	0.0291	-0.0075	0.0158	0.0044	-0.0195	-0.0390	-0.0536	0.0282	0.0083	0.0698	0.0269	-0.0092	0.0301	-0.0027	0.0000	0.0073	0.0268	0.0786	-0.0077	0.0243	-0.0341	0.0471	0.0309	0.0756	0.0659	-0.0285	-0.0358	0.0718	0.0134	-0.0121	-0.0018	0.0395	0.0114
21	0.0149	-0.0872	-0.0134	-0.0039	0.0386	-0.0284	-0.0312	-0.0117	0.0106	-0.0732	-0.0667	-0.0316	-0.0138	-0.0408	0.0153	-0.0034	-0.0230	-0.0141	-0.0673	-0.0773	0.0000	-0.0460	0.0229	0.0088	-0.0289	-0.0543	-0.0078	0.0286	0.0640	0.0259	-0.0252	-0.0351	0.0038	-0.0018	-0.0382	-0.0068	-0.0323	-0.0552
22	0.0796	0.0066	0.0507	0.0436	0.0449	-0.0254	-0.0355	0.0268	0.0619	-0.0173	-0.0665	-0.0543	0.0883	-0.0039	0.0681	-0.0384	-0.0085	-0.0048	-0.0304	-0.0268	0.0460	0.0000	0.0226	0.0058	0.0039	-0.0565	0.0156	0.0121	0.0672	0.0240	-0.0391	-0.0395	0.0557	-0.0196	-0.0321	0.0231	-0.0259	-0.0132
23	0.0210	-0.0717	-0.0275	-0.0449	-0.0120	-0.0142	-0.0299	-0.0430	-0.0268	-0.1167	-0.0776	-0.0757	-0.0223	-0.0926	0.0145	-0.0248	-0.0590	0.0118	-0.0695	-0.0786	-0.0229	-0.0226	0.0000	-0.0429	-0.0465	-0.0644	-0.0028	-0.0562	0.0317	-0.0275	-0.0837	-0.0630	-0.0318	-0.0495	-0.0787	-0.0404	-0.0685	-0.0807
24	0.0374	-0.0343	0.0203	0.0147	0.1211	0.0131	0.0020	0.0216	-0.0068	-0.0647	-0.0387	-0.0681	0.0049	-0.0195	0.0405	-0.0130	0.0131	0.0219	-0.0100	0.0077	-0.0088	-0.0058	0.0429	0.0000	0.0031	-0.0201	0.0574	0.0292	0.0630	0.0672	-0.0235	-0.0073	-0.0003	-0.0106	-0.0387	-0.0184	-0.0021	-0.0126
25	0.0495	-0.0345	0.0518	-0.0011	0.0559	0.0190	-0.0070	0.0361	0.0089	-0.0466	-0.0273	-0.0348	0.0247	-0.0129	0.0511	-0.0095	0.0041	0.0104	-0.0202	-0.0243	0.0299	-0.0039	0.0465	-0.0031	0.0000	-0.0171	0.0553	0.0494	0.1354	0.0432	-0.0075	-0.0086	0.0262	-0.0184	-0.0005	0.0008	-0.0003	-0.0242
26	0.0387	-0.0191	0.0220	0.0045	0.1025	0.0071	0.0053	0.0162	0.0197	-0.0059	-0.0168	-0.0184	0.0466	0.0537	0.0567	-0.0063	0.0188	0.0295	0.0181	0.0341	0.0343	0.0365	0.0644	0.0201	0.0171	0.0000	0.0537	0.0808	0.0844	0.0856	-0.0100	-0.0089	0.0779	0.0169	-0.0540	0.0156	0.0263	0.0027
27	-0.0023	-0.0970	-0.0135	-0.0249	0.0464	-0.0105	-0.0615	-0.0541	-0.0273	-0.1134	-0.0965	-0.0837	-0.0232	-0.0045	0.0514	-0.0304	-0.0287	-0.0379	-0.0444	-0.0471	0.0078	-0.0156	0.0028	-0.0574	-0.0353	-0.0537	0.0000	0.0137	0.0359	-0.0075	0.0723	-0.0659	0.0077	-0.0466	-0.0504	-0.0395	-0.0400	-0.0247
28	0.0232	-0.0730	0.0472	-0.0857	0.0512	-0.0382	-0.0673	-0.0371	-0.0483	-0.0714	-0.0586	-0.0648	-0.0698	-0.0035	0.0205	-0.0479	-0.0817	-0.0260	-0.0693	-0.0509	-0.0286	-0.0121	0.0362	-0.0292	-0.0094	-0.0808	-0.0137	0.0000	-0.0049	-0.0357	-0.0825	-0.0697	-0.0349	-0.0440	-0.0466	-0.0601	-0.0300	-0.0827
29	-0.0568	-0.0803	-0.0533	-0.0880	-0.0167	-0.0420	-0.1489	-0.0663	-0.0950	-0.0993	-0.1069	-0.1001	-0.0673	-0.0877	-0.0127	-0.0784	-0.0885	-0.0247	-0.1460	-0.0756	-0.0640	-0.0672	-0.0137	-0.0630	-0.1354	-0.0944	-0.0359	0.0049	0.0000	-0.0477	-0.1393	-0.1095	-0.0252	-0.0778	-0.1220	-0.0948	-0.1252	-0.0641
30	-0.0040	-0.0620	0.0387	-0.0410	0.0344	-0.0197	-0.0375	-0.0514	-0.0398	-0.1129	-0.0572	-0.0853	-0.0573	-0.0262	0.0491	-0.0373	-0.0653	-0.0194	-0.0235	-0.0639	-0.0259	-0.0240	0.0275	-0.0672	-0.0432	-0.0836	0.0075	0.0357	0.0477	0.0000	-0.0993	-0.0481	-0.0534	-0.0166	-0.0397	-0.0450	-0.0249	-0.0542
31	0.0544	-0.0160	0.0610	-0.0038	0.0744	0.0077	0.0238	0.0524	0.0648	-0.0071	0.0009	-0.0146	0.0563	0.0526	0.0796	0.0039	0.0306	0.0479	-0.0216	0.0283	0.0252	0.0391	0.0837	0.0235	0.0075	0.0100	0.0723	0.0825	0.1393	0.0993	0.0000	0.0030	0.0257	0.0225	0.0123	0.0276	0.0387	0.0379
32	0.0869	0.0032	0.0709	0.0083	0.0643	0.0278	0.0046	0.0317	0.0656	-0.0285	-0.0337	-0.0288	0.1047	0.0268	0.0896	0.0027	0.0265	0.0334	-0.0119	0.0358	0.0351	0.0395	0.0650	0.0073	0.0086	0.0089	0.0659	0.0697	0.1095	0.0481	-0.0030	0.0000	0.0653	0.0093	-0.0056	0.0258	0.0379	0.0105
33	0.0210	-0.0877	0.0071	-0.0220	0.0016	-0.0693	-0.1052	-0.0564	-0.0031	-0.0859	-0.0527	-0.0530	0.0049	-0.0298	0.0294	-0.0794	-0.0098	-0.0464	-0.0746	-0.0249	-0.0178	0.0018	0.0196	0.0495	0.0106	0.0184	-0.0169	0.0466	0.0440	0.0778	0.0166	-0.0225	-0.0093	0.0737	0.0000	0.0073	-0.0207	-0.0235
34	0.0419	-0.0352	-0.0077	-0.0091	0.0505	0.0070	-0.0184	0.0245	0.0236	-0.0301	-0.0344	-0.0319	0.0353	-0.0044	0.0290	0.0197	-0.0009	0.0245	-0.0249	-0.0178	0.0018	0.0196	0.0495	0.0106	0.0184	-0.0169	0.0466	0.0440	0.0778	0.0166	-0.0225	-0.0093	0.0737	0.0000	0.0073	-0.0207	-0.0233	-0.0235
35	0.0740	-0.0231	0.0616	0.0298	0.0494	-0.0002	-0.0189	0.0599	0.0424	-0.0224	-0.0305	-0.0190	0.0336	0.0074	0.0749	-0.0020	0.0115	0.0087	-0.0367	-0.0134	0.0382	0.0321	0.0787	0.0387	0.0005	0.0340	0.0504	0.0466	0.1220	0.0397	-0.0123	0.0056	0.0424	-0.0073	0.0000	0.0101	-0.0234	0.0173
36	0.0725	-0.0018	0.0279	-0.0050	0.0263	-0.0110	-0.0233	0.0222	0.0199	-0.0109	-0.0621	-0.0421	0.0272	-0.0560	0.0734	0.0024	0.0019	-0.0163	-0.0394	0.0121	0.0068	-0.0231	0.0404	0.0184	-0.0008	-0.0156	0.0395	0.0601	0.0948	0.0459	-0.0276	-0.0258	0.0318	0.0207	-0.0301	0.0000	-0.0286	-0.0241
37	0.0587	-0.0151	0.0309	0.0032	0.0716	-0.0131	-0.0071	0.0131	0.0354	-0.0456	0.0126	-0.0221	0.0340	0.0363	0.0637	-0.0066	0.0162	0.0183	-0.0395	0.0018	0.0323	0.0259	0.0685	0.0021	0.0003	-0.0263	0.0400	0.0300	0.1252	0.0249	-0.0387	-0.0379	0.0470	0.0233	0.0234	0.0286	0.0000	0.0096
38	0.0958	0.0128	0.0610	0.0366	0.0383	0.0259	-0.0003	0.0283	0.0271	0.0093	-0.0384	-0.0461	0.0435	-0.0142	0.0606	0.0077	-0.0057	0.0445	-0.0114	0.0035	0.0552	0.0132	0.0807	0.0126	0.0242	-0.0127	0.0247	0.0827	0.0641	0.0542	-0.0379	-0.0105	0.0769	0.0255	-0.0173	0.0241	-0.0096	0.0000

Table B.2: Magnitudes of differences between δ_J of each dataset.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	0.0000	0.0702	0.0480	0.0306	0.0321	0.0636	0.0752	0.0335	0.0486	0.0418	0.0963	0.1010	0.0464	0.0552	0.0144	0.0335	0.0529	0.0695	0.0674	0.0699	0.0149	0.0796	0.0210	0.0374	0.0495	0.0387	0.0023	0.0232	0.0568	0.0040	0.0544	0.0869	0.0210	0.0419	0.0740	0.0725	0.0587	0.0958
2	0.0702	0.0000	0.0884	0.0253	0.0771	0.0217	0.0083	0.0836	0.0091	0.0293	0.0133	0.0060	0.0151	0.0504	0.1319	0.0322	0.0257	0.0319	0.0196	0.0147	0.0872	0.0066	0.0717	0.0343	0.0345	0.0191	0.0970	0.0730	0.0803	0.0620	0.0160	0.0032	0.0877	0.0352	0.0231	0.0018	0.0151	0.0128
3	0.0480	0.0884	0.0000	0.0159	0.0236	0.0665	0.0813	0.0236	0.0132	0.0731	0.0469	0.0456	0.0161	0.0129	0.0304	0.0186	0.0381	0.0446	0.0804	0.0797	0.0134	0.0507	0.0273	0.0203	0.0518	0.0220	0.0155	0.0472	0.0533	0.0587	0.0610	0.0709	0.0071	0.0077	0.0616	0.0279	0.0509	0.0610
4	0.0306	0.0253	0.0159	0.0000	0.0726	0.0096	0.0076	0.0226	0.0439	0.0407	0.0145	0.0200	0.0615	0.0150	0.0404	0.0100	0.0093	0.0422	0.0097	0.0469	0.0039	0.0436	0.0449	0.0147	0.0011	0.0045	0.0249	0.0857	0.0880	0.0410	0.0038	0.0083	0.0220	0.0091	0.0298	0.0050	0.0032	0.0366
5	0.0321	0.0771	0.0236	0.0726	0.0000	0.0601	0.1073	0.0633	0.0286	0.0836	0.1071	0.1155	0.0321	0.0026	0.0189	0.0747	0.0465	0.0303	0.0707	0.0802	0.0386	0.0449	0.0120	0.1211	0.0559	0.1025	0.0464	0.0512	0.0167	0.0344	0.0744	0.0643	0.0016	0.0505	0.0494	0.0263	0.0716	0.0583
6	0.0636	0.0217	0.0665	0.0096	0.0601	0.0000	0.0294	0.0164	0.0525	0.0202	0.0706	0.0229	0.0595	0.0169	0.0855	0.0019	0.0148	0.0138	0.0196	0.0291	0.0284	0.0254	0.0142	0.0131	0.0190	0.0071	0.0105	0.0382	0.0420	0.0197	0.0077	0.0278	0.0693	0.0070	0.0002	0.0110	0.0131	0.0259
7	0.0752	0.0083	0.0813	0.0076	0.1073	0.0294	0.0000	0.0684	0.0567	0.0391	0.0063	0.0157	0.0650	0.0359	0.0713	0.0158	0.0279	0.0522	0.0012	0.0075	0.0312	0.0355	0.0599	0.0020	0.0070	0.0053	0.0615	0.0673	0.1489	0.0375	0.0258	0.0046	0.1052	0.0184	0.0189	0.0233	0.0071	0.0003
8	0.0335	0.0836	0.0236	0.0226	0.0633	0.0164	0.0684	0.0000	0.0022	0.0276	0.0644	0.0702	0.0103	0.0519	0.0282	0.0340	0.0360	0.0019	0.0419	0.0158	0.0117	0.0268	0.0430	0.0216	0.0261	0.0162	0.0541	0.0571	0.0663	0.0514	0.0524	0.0317	0.0364	0.0245	0.0599	0.0222	0.0131	0.0283
9	0.0486	0.0091	0.0132	0.0439	0.0286	0.0525	0.0567	0.0022	0.0000	0.0406	0.0088	0.0289	0.0162	0.0283	0.0439	0.0079	0.0020	0.0063	0.0445	0.0044	0.0106	0.0619	0.0268	0.0068	0.0089	0.0197	0.0273	0.0483	0.0950	0.0398	0.0648	0.0656	0.0031	0.0226	0.0424	0.0199	0.0354	0.0271
10	0.0418	0.0293	0.0731	0.0407	0.0836	0.0202	0.0391	0.0276	0.0406	0.0000	0.0025	0.0263	0.0570	0.0162	0.0433	0.0317	0.0019	0.0155	0.0337	0.0305	0.0732	0.0173	0.1167	0.0647	0.0466	0.0059	0.1134	0.0714	0.0993	0.1129	0.0071	0.0285	0.0569	0.0301	0.0224	0.0109	0.0456	0.0015
11	0.0963	0.0133	0.0469	0.0145	0.1071	0.0706	0.0063	0.0644	0.0068	0.0025	0.0000	0.0067	0.0203	0.0396	0.0997	0.0538	0.0318	0.0687	0.0104	0.0390	0.0667	0.0605	0.0776	0.0387	0.0273	0.0168	0.0865	0.0586	0.1669	0.0572	0.0009	0.0337	0.0527	0.0344	0.0305	0.0621	0.0126	0.0384
12	0.1010	0.0060	0.0456	0.0200	0.1155	0.0229	0.0157	0.0702	0.0289	0.0263	0.0067	0.0000	0.0419	0.0116	0.0773	0.0670	0.0279	0.0431	0.0100	0.0536	0.0316	0.0543	0.0737	0.0681	0.0348	0.0184	0.0837	0.0648	0.1901	0.0853	0.0146	0.0288	0.0530	0.0319	0.0190	0.0421	0.0221	0.0461
13	0.0464	0.0151	0.0161	0.0615	0.0321	0.0595	0.0650	0.0103	0.0162	0.0570	0.0203	0.0419	0.0000	0.0390	0.0449	0.0300	0.0146	0.0094	0.0682	0.0282	0.0138	0.0883	0.0223	0.0049	0.0232	0.0698	0.0673	0.0573	0.0563	0.1047	0.0049	0.0353	0.0336	0.0272	0.0340	0.0455
14	0.0552	0.0504	0.0129	0.0150	0.0026	0.0169	0.0359	0.0519	0.0283	0.0116	0.0396	0.0000	0.0390	0.0000	0.0841	0.0138	0.0639	0.0250	0.0476	0.0083	0.0408	0.0039	0.0926	0.0195	0.0129	0.0537	0.0045	0.0035	0.0877	0.0262	0.0526	0.0268	0.0298	0.0044	0.0074	0.0560	0.0363	0.0142
15	0.0144	0.1319	0.0304	0.0404	0.0189	0.0855	0.0713	0.0282	0.0843	0.0997	0.0773	0.0449	0.0843	0.0000	0.0791	0.0813	0.0720	0.0831	0.0203	0.0159	0.0662	0.0000	0.0027	0.0673	0.0304	0.0695	0.0100	0.0202	0.0181	0.0235	0.0216	0.0119	0.0746	0.0249	0.0367	0.0394	0.0395	0.0114
16	0.0335	0.0322	0.0186	0.0100	0.0747	0.0019	0.0158	0.0340	0.0079	0.0317	0.0538	0.0670	0.0300	0.0138	0.0791	0.0000	0.0105	0.0183	0.0203	0.0269	0.0034	0.0384	0.0248	0.0130	0.0095	0.0063	0.0304	0.0479	0.0784	0.0373	0.0039	0.0027	0.0794	0.0197	0.0020	0.0024	0.0066	0.0077
17	0.0529	0.0257	0.0381	0.0093	0.0465	0.0148	0.0279	0.0360	0.0020	0.0019	0.0318	0.0279	0.0146	0.0639	0.0813	0.0105	0.0000	0.0307	0.0159	0.0092	0.0250	0.0085	0.0590	0.0131	0.0041	0.0188	0.0287	0.0817	0.0885	0.0653	0.0306	0.0266	0.0098	0.0009	0.0115	0.0019	0.0162	0.0057
18	0.0695	0.0319	0.0446	0.0422	0.0303	0.0138	0.0522	0.0019	0.0063	0.0155	0.0687	0.0431	0.0094	0.0250	0.0720	0.0183	0.0307	0.0000	0.0662	0.0301	0.0141	0.0948	0.0118	0.0219	0.0104	0.0293	0.0379	0.0260	0.0247	0.0194	0.0479	0.0334	0.0464	0.0245	0.0087	0.0163	0.0183	0.0445
19	0.0674	0.0196	0.0804	0.0097	0.0707	0.0196	0.0012	0.0419	0.0445	0.0337	0.0104	0.0100	0.0682	0.0476	0.0831	0.0203	0.0159	0.0662	0.0000	0.0027	0.0673	0.0304	0.0695	0.0100	0.0202	0.0181	0.0444	0.0693	0.1460	0.0235	0.0216	0.0119	0.0746	0.0249	0.0367	0.0394	0.0395	0.0114
20	0.0699	0.0147	0.0797	0.0469	0.0802	0.0291	0.0075	0.0158	0.0044	0.0305	0.0390	0.0536	0.0282	0.0083	0.0698	0.0269	0.0092	0.0301	0.0027	0.0000	0.0773	0.0268	0.0786	0.0077	0.0243	0.0341	0.0471	0.0509	0.0756	0.0639	0.0285	0.0358	0.0718	0.0178	0.0134	0.0121	0.0018	0.0035
21	0.0149	0.0872	0.0134	0.0039	0.0386	0.0284	0.0312	0.0117	0.0106	0.0732	0.0667	0.0316	0.0138	0.0408	0.0153	0.0034	0.0250	0.0141	0.0673	0.0773	0.0000	0.0460	0.0229	0.0088	0.0299	0.0343	0.0078	0.0286	0.0640	0.0259	0.0252	0.0351	0.0038	0.0018	0.0382	0.0068	0.0323	0.0552
22	0.0796	0.0066	0.0507	0.0436	0.0449	0.0254	0.0355	0.0268	0.0619	0.0173	0.0605	0.0543	0.0883	0.0039	0.0681	0.0384	0.0085	0.0048	0.0304	0.0268	0.0460	0.0000	0.0226	0.0058	0.0039	0.0365	0.0156	0.0121	0.0672	0.0240	0.0391	0.0395	0.0557	0.0196	0.0321	0.0231	0.0259	0.0132
23	0.0210	0.0717	0.0273	0.0449	0.0120	0.0142	0.0599	0.0430	0.0268	0.1167	0.0776	0.0737	0.0223	0.0926	0.0115	0.0248	0.0390	0.0118	0.0695	0.0786	0.0229	0.0226	0.0000	0.0429	0.0465	0.0644	0.0028	0.0362	0.0317	0.0275	0.0837	0.0630	0.0318	0.0495	0.0787	0.0404	0.0685	0.0807
24	0.0374	0.0343	0.0203	0.0147	0.1211	0.0131	0.0020	0.0216	0.0068	0.0647	0.0387	0.0681	0.0049	0.0195	0.0405	0.0130	0.0131	0.0219	0.0100	0.0077	0.0088	0.0058	0.0429	0.0000	0.0031	0.0201	0.0574	0.0292	0.0630	0.0672	0.0235	0.0073	0.0003	0.0106	0.0387	0.0184	0.0021	0.0126
25	0.0495	0.0345	0.0518	0.0011	0.0559	0.0190	0.0070	0.0261	0.0089	0.0466	0.0273	0.0348	0.0247	0.0129	0.0511	0.0095	0.0041	0.0104	0.0202	0.0243	0.0299	0.0039	0.0465	0.0031	0.0000	0.0171	0.0553	0.0494	0.1354	0.0432	0.0075	0.0086	0.0262	0.0184	0.0005	0.0008	0.0003	0.0242
26	0.0387	0.0191	0.0220	0.0045	0.1025	0.0071	0.0053	0.0162	0.0197	0.0959	0.0168	0.0184	0.0466	0.0537	0.0367	0.0063	0.0188	0.0293	0.0181	0.0341	0.0343	0.0365	0.0644	0.0201	0.0171	0.0000	0.0537	0.0808	0.0944	0.0836	0.0100	0.0089	0.0779	0.0169	0.0340	0.0156	0.0263	0.0027
27	0.0023	0.0970	0.0155	0.0249	0.0464	0.0105	0.0615	0.0541	0.0273	0.1134	0.0865	0.0837	0.0232	0.0045	0.0514	0.0304	0.0287	0.0379	0.0444	0.0471	0.0978	0.0156	0.0028	0.0574	0.0553	0.0537	0.0000	0.0137	0.0359	0.0075	0.0723	0.0659	0.0077	0.0466	0.0504	0.0395	0.0400	0.0247
28	0.0232	0.0730	0.0472	0.0857	0.0512	0.0382	0.0673	0.0571	0.0483	0.0714	0.0586	0.0648	0.0698	0.0035	0.0205	0.0479	0.0817	0.0260	0.0693	0.0509	0.0286	0.0121	0.0362	0.0292	0.0494	0.0808	0.0137	0.0000	0.0049	0.0357	0.0825	0.0697	0.0349	0.0440	0.0466	0.0601	0.0300	0.0827
29	0.0568	0.0803	0.0533	0.0880	0.0167	0.0420	0.1489	0.0663	0.0950	0.0993	0.1669	0.1001	0.0673	0.0877	0.0127	0.0784	0.0885	0.0247	0.1460	0.0756	0.0640	0.0672	0.0317	0.0630	0.1354	0.0944	0.0808	0.0137	0.0000	0.0049	0.0000	0.0477	0.1393	0.1095	0.0252	0.0778	0.1220	0.0948	0.1252	0.0641
30	0.0040	0.0620	0.0587	0.0410	0.0344	0.0197	0.0375	0.0514	0.0398	0.1129	0.0572	0.0853	0.0573	0.0262	0.0491	0.0373	0.0653	0.0194	0.0235	0.0639	0.0259	0.0240	0.0275	0.0672	0.0432	0.0836	0.0075	0.0357	0.0477	0.0000	0.0993	0.0481	0.0534	0.0166	0.0397	0.0450	0.0249	0.0542
31	0.0544	0.0160	0.0610	0.0038	0.0744	0.0077	0.0258	0.0528	0.0648	0.0971	0.0009	0.0146	0.0563	0.0526	0.0796	0.0939	0.0306	0.0479	0.0216	0.0285	0.0252	0.0391	0.0837	0.0235	0.0075	0.0100	0.0723	0.0825	0.1393	0.0993	0.0000	0.0930	0.0257	0.0225	0.0123	0.0276	0.0387	0.0379
32	0.0869	0.0032	0.0709	0.0083	0.0643	0.0278	0.0046	0.0317	0.0656	0.0285	0.0337	0.0288	0.1047	0.0268	0.0896	0.0027	0.0266	0.0334	0.0119	0.0358	0.0351	0.0395	0.0630	0.0073	0.0086	0.0089	0.0659	0.0697	0.1095	0.0481	0.0030	0.0000	0.0653	0.0093	0.0056	0.0258	0.0379	0.0105
33	0.0210	0.0877	0.0071	0.0220	0.0016	0.0693	0.1052	0.0364	0.0031	0.0509	0.0527	0.0530	0.0049	0.0298	0.0294	0.0794	0.0098	0.0464	0.0746	0.0718	0.0038	0.0557	0.0318	0.0003	0.0262	0.0779	0.0977	0.0349	0.0252	0.0534	0.0257	0.0653	0.0000	0.0737	0.0424	0.0318	0.0470	0.0769
34	0.0419	0.0352	0.0077	0.0091	0.0505	0.0070	0.0184	0.0245	0.0236	0.0301	0.0344	0.0319	0.0353	0.0044	0.0290	0.0197	0.0069	0.0245	0.0249	0.0178	0.0018	0.0196	0.0495	0.0106	0.0184	0.0169	0.0466	0.0440	0.0778	0.0166	0.0225	0.0093	0.0737	0.0000	0.0073	0.0207	0.0233	0.0255
35	0.0740	0.0231	0.0616	0.0298	0.0494	0.0302	0.0189	0.0599	0.0424	0.0234	0.0305	0.0190	0.0336	0.0074	0.0749	0.0020	0.0115	0.0087	0.0367	0.0134	0.0382	0.0321	0.0787	0.0387	0.0005	0.0340	0.0169	0.0466	0.1220	0.0397	0.0123	0.0056	0.0424	0.0073	0.0000	0.0301	0.0234	0.0173
36	0.0725	0.0018	0.0279	0.0050	0.0263	0.0110	0.0233	0.0222	0.0199	0.0109	0.0621	0.0421	0.0272	0.0560	0.0734	0.0024	0.0019	0.0163	0.0394	0.0121	0.0068	0.0231	0.0404	0.0184	0.0008	0.0156	0.0395	0.0601	0.0948	0.0450	0.0276	0.0258	0.0318	0.0207	0.0301	0.0000	0.0286	0.0241
37	0.0587	0.0151	0.0509	0.0032	0.0716	0.0131	0.0071	0.0131	0.0354	0.0456	0.0126	0.0221	0.0340	0.0363	0.0637	0.0066	0.0162	0.0183	0.0395	0.0018	0.0323	0.0259	0.0685	0.0021	0.0003	0.0263	0.0400	0.0300	0.1252	0.0249	0.0387	0.0379	0.0470	0.0233	0.0234	0.0286	0.0000	0.0096
38	0.0958	0.0128	0.0610	0.0366	0.0583	0.0259	0.0003	0.0283	0.0271	0.0015	0.0384	0.0461	0.0455	0.0142	0.0806	0.0077	0.0057	0.0445	0.0114	0.0035	0.0552	0.0132	0.0807	0.0126	0.0242	0.0027	0.0247	0.0827	0.0641	0.0542	0.0379	0.0105	0.0769	0.0255	0.0173	0.0241	0.0096	0.0000

Table B.3: Differences between δ_S of each dataset.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	0.0000	-0.0355	0.0055	-0.0263	0.0249	-0.0471	-0.0643	-0.0054	-0.0291	-0.0254	-0.0681	-0.0769	-0.0382	-0.0335	0.0182	-0.0214	-0.0339	-0.0733	-0.0384	-0.0631	-0.0352	-0.0820	-0.0208	-0.0646	-0.0453	-0.0343	0.0296	0.0232	0.0769	0.0411	-0.0357	-0.0530	-0.0216	-0.0378	-0.0479	-0.0643	-0.0266	-0.0660
2	0.0355	0.0000	0.0847	0.0134	0.0416	-0.0069	-0.0144	0.0629	-0.0105	0.0007	-0.0284	-0.0231	-0.0163	0.0006	0.1160	0.0106	0.0141	-0.0070	-0.0250	-0.0134	0.0418	-0.0200	0.0226	-0.0243	-0.0059	-0.0033	0.0795	0.0631	0.0678	0.0318	-0.0027	0.0106	0.0484	0.0025	0.0209	-0.0188	0.0012	-0.0232
3	-0.0055	-0.0847	0.0000	-0.0332	-0.0027	-0.0905	-0.0895	-0.0233	-0.0407	-0.0595	-0.0715	-0.0679	-0.0399	-0.0367	0.0086	-0.0156	-0.0548	-0.0977	-0.1031	-0.1003	-0.0350	-0.0942	-0.0294	-0.0549	-0.0794	-0.0476	0.0066	-0.0278	0.0618	-0.0494	-0.0601	-0.0797	-0.0162	-0.0235	-0.0754	-0.0634	-0.0532	-0.0729
4	0.0263	-0.0134	0.0332	0.0000	0.0652	-0.0032	0.0049	0.0111	0.0322	-0.0459	-0.0145	-0.0383	0.0499	-0.0137	0.0236	-0.0106	0.0147	-0.0095	-0.0297	-0.0418	-0.0296	-0.0356	0.0200	-0.0359	-0.0044	-0.0214	0.0128	0.0445	0.1025	0.0695	0.0103	0.0192	0.0159	-0.0153	-0.0260	-0.0947	-0.0130	-0.0355
5	-0.0249	-0.0416	0.0027	-0.0652	0.0000	-0.0528	-0.0890	-0.0298	-0.0211	-0.0699	-0.0901	-0.0798	-0.0773	0.0041	0.0181	-0.0402	-0.0460	-0.0454	-0.0847	-0.0739	-0.0315	-0.0456	-0.0141	-0.1193	-0.0699	-0.0713	-0.0177	-0.0406	0.0328	-0.0333	-0.0824	-0.0405	-0.0052	-0.0561	-0.0482	-0.0368	-0.0648	-0.0387
6	0.0471	0.0069	0.0905	0.0032	0.0528	0.0000	0.0065	0.0353	0.0343	-0.0132	-0.0190	-0.0255	0.0504	-0.0003	0.0767	0.0236	0.0204	-0.0114	-0.0224	-0.0053	-0.0013	-0.0019	0.0108	-0.0206	-0.0083	-0.0071	0.0481	0.0565	0.0808	0.0340	0.0084	0.0147	0.0700	-0.0053	0.0153	-0.0015	0.0261	-0.0154
7	0.0643	0.0144	0.0895	-0.0049	0.0890	-0.0065	0.0000	0.0491	0.0534	-0.0231	-0.0111	-0.0222	0.0608	0.0137	0.0869	-0.0062	0.0361	-0.0084	-0.0214	-0.0051	0.0068	-0.0048	0.0064	-0.0228	-0.0101	0.0017	0.0651	0.0607	0.1311	0.0581	-0.0236	0.0127	0.0832	-0.0039	0.0224	0.0049	-0.0072	-0.0044
8	0.0054	-0.0629	0.0233	-0.0111	0.0298	-0.0353	-0.0491	0.0000	-0.0066	-0.0526	-0.0533	-0.0615	-0.0065	-0.0495	0.0238	-0.0137	-0.0311	-0.0282	-0.0567	-0.0443	-0.0036	-0.0670	0.0006	-0.0439	-0.0385	-0.0344	0.0539	0.0440	0.0448	0.0180	-0.0271	-0.0271	0.0058	-0.0258	-0.0292	-0.0387	-0.0150	-0.0569
9	0.0291	0.0105	0.0407	-0.0322	0.0211	-0.0343	-0.0534	0.0066	0.0000	-0.0329	-0.0128	-0.0390	-0.0034	-0.0196	0.0323	-0.0161	-0.0179	-0.0325	-0.0425	-0.0336	-0.0076	-0.0790	-0.0037	-0.0324	-0.0351	-0.0322	0.0316	0.0546	0.0853	0.0579	-0.0359	-0.0523	0.0065	-0.0256	-0.0304	-0.0395	-0.0226	-0.0488
10	0.0254	-0.0007	0.0595	0.0459	0.0699	0.0132	0.0231	0.0526	0.0329	0.0000	0.0174	-0.0183	0.0282	0.0106	0.0317	0.0523	0.0123	0.0118	-0.0012	0.0038	0.0216	-0.0025	0.0624	0.0193	0.0308	0.0147	0.1249	0.0725	0.0951	0.0871	0.0205	0.0189	0.0285	0.0202	0.0241	0.0224	0.0177	0.0079
11	0.0681	0.0284	0.0715	0.0145	0.0901	0.0190	0.0111	0.0533	0.0128	0.0174	0.0000	-0.0107	0.0218	0.0281	0.0881	0.0276	0.0373	0.0155	0.0036	0.0204	0.0217	0.0225	0.0268	-0.0075	0.0132	0.0093	0.0790	0.0779	0.1434	0.0647	0.0088	0.0347	0.0392	0.0281	0.0204	0.0356	0.0122	0.0185
12	0.0769	0.0231	0.0679	0.0383	0.0798	0.0255	0.0222	0.0615	0.0390	0.0183	0.0107	0.0000	0.0283	0.0062	0.0916	0.0557	0.0505	0.0169	0.0071	0.0186	0.0287	0.0194	0.0598	0.0075	0.0274	0.0235	0.0960	0.1036	0.0806	0.0815	0.0181	0.0362	0.0283	0.0417	0.0363	0.0277	0.0254	0.0198
13	0.0382	0.0163	0.0399	-0.0499	0.0173	-0.0504	-0.0608	0.0065	0.0034	-0.0282	-0.0218	-0.0283	0.0000	-0.0156	0.0445	-0.0225	-0.0211	-0.0315	-0.0689	-0.0363	-0.0045	-0.0847	0.0005	-0.0251	-0.0478	-0.0356	0.0326	0.0576	0.0800	0.0739	-0.0341	-0.0645	0.0013	-0.0337	-0.0296	-0.0398	-0.0250	-0.0469
14	0.0335	-0.0006	0.0367	0.0137	-0.0041	0.0003	-0.0137	0.0495	0.0196	-0.0106	-0.0281	-0.0062	0.0156	0.0000	0.0891	0.0199	-0.0294	0.0011	-0.0342	-0.0258	0.0116	-0.0140	0.0281	-0.0100	0.0024	-0.0217	0.0473	0.0386	0.1179	0.0349	-0.0222	0.0058	0.0012	0.0133	0.0150	0.0291	-0.0021	-0.0083
15	-0.0182	-0.1160	-0.0086	-0.0226	-0.0181	-0.0767	-0.0869	-0.0238	-0.0323	-0.0699	-0.0901	-0.0917	-0.0445	-0.0891	0.0000	-0.0653	-0.0771	-0.0716	-0.0829	-0.0799	-0.0409	-0.0653	-0.0771	-0.0815	-0.0696	-0.0467	-0.0372	-0.0032	0.0223	-0.0165	-0.0607	-0.0602	-0.0327	-0.0244	-0.0589	-0.0643	-0.0498	-0.0735
16	0.0214	-0.0106	0.0156	0.0106	0.0402	-0.0236	0.0062	0.0137	0.0161	-0.0523	-0.0276	-0.0557	0.0225	-0.0199	0.0653	0.0000	0.0053	-0.0058	-0.0249	-0.0125	-0.0292	-0.0268	0.0089	-0.0164	-0.0129	-0.0071	0.0474	0.0397	0.0825	0.0350	0.0009	0.0111	0.0623	-0.0158	0.0103	-0.0170	0.0155	-0.0246
17	0.0339	-0.0141	0.0548	-0.0147	0.0460	-0.0204	-0.0361	0.0311	0.0179	-0.0123	-0.0373	-0.0505	0.0211	0.0294	0.0771	-0.0053	0.0000	-0.0144	-0.0231	-0.0182	0.0074	-0.0177	0.0131	-0.0388	-0.0109	-0.0233	0.0283	0.0599	0.0794	0.0641	-0.0150	-0.0062	0.0352	-0.0096	-0.0089	-0.0361	-0.0085	-0.0074
18	0.0733	0.0070	0.0977	0.0095	0.0454	0.0114	0.0084	0.0282	0.0325	-0.0118	-0.0155	-0.0169	0.0315	-0.0011	0.0736	0.0058	0.0144	0.0000	-0.0141	-0.0036	0.0221	-0.0013	0.0089	-0.0183	-0.0061	-0.0033	0.0540	0.0630	0.0857	0.0505	0.0092	0.0209	0.0613	0.0101	0.0173	0.0039	0.0341	-0.0159
19	0.0584	0.0250	0.1031	0.0297	0.0847	0.0224	0.0214	0.0567	0.0425	0.0012	-0.0036	-0.0071	0.0689	0.0342	0.0829	0.0249	0.0231	0.0141	0.0000	0.0059	0.0599	0.0062	0.0373	-0.0073	0.0194	0.0013	0.0563	0.0748	0.1372	0.0663	0.0283	0.0232	0.0844	0.0219	0.0398	0.0181	0.0328	0.0174
20	0.0631	0.0134	0.1003	0.0418	0.0739	0.0053	0.0051	0.0443	0.0336	-0.0038	-0.0204	-0.0186	0.0363	0.0258	0.0799	0.0125	0.0182	0.0036	-0.0059	0.0000	0.0490	0.0097	0.0462	-0.0131	0.0161	0.0040	0.0614	0.0765	0.0808	0.0808	0.0137	-0.0000	0.0699	0.0001	0.0228	0.0083	0.0201	0.0203
21	0.0352	-0.0418	0.0350	0.0296	0.0315	0.0013	-0.0068	0.0036	0.0076	-0.0216	-0.0217	-0.0287	0.0045	-0.0116	0.0409	0.0292	-0.0074	-0.0221	-0.0599	-0.0490	0.0000	-0.0545	0.0142	-0.0187	-0.0276	-0.0140	0.0293	0.0417	0.1007	0.0646	-0.0133	0.0140	0.0220	0.0228	-0.0028	0.0030	-0.0157	-0.0367
22	0.0820	0.0200	0.0942	0.0356	0.0456	0.0019	0.0048	0.0670	0.0790	0.0025	-0.0225	-0.0194	0.0847	0.0140	0.0779	0.0268	0.0177	0.0013	-0.0062	-0.0097	0.0545	0.0000	0.0027	-0.0104	-0.0108	0.0021	0.0528	0.0573	0.0968	0.0694	0.0178	0.0143	0.0573	0.0171	0.0207	0.0241	0.0241	0.0109
23	0.0208	-0.0226	0.0294	-0.0200	0.0141	-0.0108	-0.0064	-0.0006	0.0037	-0.0624	-0.0268	-0.0598	-0.0005	-0.0281	0.0355	-0.0089	-0.0131	-0.0089	-0.0373	-0.0462	-0.0142	-0.0027	0.0000	-0.0546	-0.0374	-0.0468	0.0314	0.0229	0.0696	0.0227	-0.0382	-0.0042	0.0061	-0.0119	-0.0410	-0.0225	-0.0340	-0.0561
24	0.0646	0.0243	0.0549	0.0359	0.1193	0.0206	0.0228	0.0439	0.0324	-0.0193	0.0075	-0.0075	0.0251	0.0100	0.0815	0.0164	0.0388	0.0183	0.0073	0.0131	0.0187	0.0104	0.0546	0.0000	0.0216	0.0155	0.0885	0.0716	0.0728	0.0805	0.0108	0.0358	0.0254	0.0228	0.0293	0.0190	0.0199	0.0096
25	0.0453	0.0059	0.0794	0.0044	0.0669	0.0083	0.0101	0.0385	0.0351	-0.0308	-0.0132	-0.0274	0.0478	-0.0024	0.0696	0.0129	0.0109	0.0061	-0.0194	-0.0161	0.0276	0.0108	0.0374	-0.0216	0.0000	-0.0161	0.0578	0.0568	0.1251	0.0480	0.0043	0.0137	0.0509	0.0080	0.0173	0.0051	0.0106	0.0077
26	0.0343	0.0033	0.0476	0.0214	0.0713	0.0071	-0.0017	0.0344	0.0322	-0.0147	-0.0093	-0.0235	0.0356	0.0217	0.0467	0.0071	0.0233	0.0033	-0.0013	-0.0040	0.0140	-0.0021	0.0468	-0.0155	0.0161	0.0000	0.0614	0.0568	0.0807	0.0772	0.0033	0.0177	0.0630	0.0150	0.0066	0.0033	0.0165	0.0034
27	-0.0296	-0.0795	-0.0066	-0.0128	0.0177	-0.0481	-0.0651	-0.0539	-0.0316	-0.1249	-0.0790	-0.0960	-0.0326	-0.0473	0.0372	-0.0474	-0.0283	-0.0540	-0.0563	-0.0614	-0.0293	-0.0528	-0.0314	-0.0885	-0.0578	-0.0614	0.0000	0.0251	0.0251	-0.0019	-0.0566	-0.0383	0.0038	-0.0466	-0.0397	-0.0445	-0.0438	-0.0591
28	-0.0232	-0.0631	0.0278	-0.0445	0.0406	-0.0565	-0.0607	-0.0440	-0.0546	-0.0725	-0.0779	-0.1036	-0.0576	-0.0386	0.0032	-0.0397	-0.0599	-0.0630	-0.0748	-0.0765	-0.0417	-0.0573	-0.0229	-0.0716	-0.0568	-0.0568	-0.0251	0.0000	-0.0213	-0.0162	-0.0796	-0.0513	-0.0260	-0.0541	-0.0531	-0.0683	-0.0596	-0.0815
29	-0.0769	-0.0678	-0.0618	-0.1025	-0.0328	-0.0808	-0.1311	-0.0448	-0.0853	-0.0951	-0.1434	-0.0806	-0.0800	-0.1179	-0.0223	-0.0825	-0.0794	-0.0857	-0.1372	-0.0808	-0.1007	-0.0968	-0.0696	-0.0728	-0.1251	-0.0807	-0.0251	0.0213	0.0000	-0.0348	-0.1096	-0.0710	-0.0377	-0.0849	-0.1096	-0.0912	-0.0980	-0.0850
30	-0.0411	-0.0318	0.0494	-0.0695	0.0333	-0.0340	-0.0581	-0.0180	-0.0579	-0.0871	-0.0647	-0.0815	-0.0739	-0.0349	0.0165	-0.0350	-0.0641	-0.0505	-0.0663	-0.0808	-0.0646	-0.0694	-0.0227	-0.0805	-0.0480	-0.0772	0.0019	0.0162	0.0348	0.0000	-0.0836	-0.0382	-0.0423	-0.0330	-0.0531	-0.0529	-0.0541	-0.0746
31	0.0357	0.0027	0.0601	-0.0103	0.0824	-0.0084	0.0226	0.0271	0.0359	-0.0205	-0.0088	-0.0181	0.0341	0.0222	0.0607	-0.0009	0.0150	-0.0092	-0.0283	-0.0137	0.0133	-0.0178	0.0382	-0.0108	-0.0043	-0.0033	0.0566	0.0796	0.1096	0.0836	0.0000	-0.0077	0.0387	-0.0046	-0.0028	0.0012	0.0130	-0.0020
32	0.0530	-0.0106	0.0797	-0.0192	0.0405	-0.0147	-0.0127	0.0271	0.0523	-0.0189	-0.0347	-0.0362	0.0645	-0.0058	0.0602	-0.0111	0.0062	-0.0209	-0.0232	0.0000	-0.0140	-0.0143	0.0042	-0.0358	-0.0137	-0.0177	0.0383	0.0513	0.0710	0.0382	0.0077	0.0000	0.0427	-0.0127	-0.0171	-0.0152	0.0252	-0.0232
33	0.0216	-0.0484	0.0162	-0.0159	0.0052	-0.0700	-0.0832	-0.0038	-0.0065	-0.0285	-0.0392	-0.0283	-0.0013	-0.0012	0.0327	-0.0623	-0.0352	-0.0613	-0.0844	-0.0699	-0.0220	-0.0573	-0.0061	-0.0254	-0.0509	-0.0630	-0.0038	0.0260	0.0377	0.0423	-0.0387	-0.0427	0.0000	-0.0702	-0.0379	-0.0433	-0.0582	-0.0487
34	0.0378	-0.0025	0.0235	0.0153	0.0561	0.0053	0.0039	0.0258	0.0256	-0.0202	-0.0281	-0.0417	0.0337	-0.0133	0.0244	0.0158	0.0096	-0.0101	-0.0219	-0.0001	-0.0228	-0.0171	0.0119	-0.0228	-0.0080	-0.0150	0.0466	0.0541	0.0849	0.0330	0.0046	0.0127	0.0702	0.0000	0.0100	-0.0101	-0.0088	-0.0205
35	0.0479	-0.0209	0.0754	0.0260	0.0482	-0.0153	-0.0224	0.0392	0.0304	-0.0241	-0.0204	-0.0363	0.0296	-0.0150	0.0589	-0.0103	0.0089	-0.0173	-0.0398	-0.0228	0.0028	-0.0207	0.0410	-0.0293	-0.0173	-0.0066	0.0397	0.0531	0.1096	0.0531	0.0028	0.0171	0.0379	-0.0100	0.0000	0.0015	-0.0200	-0.0168
36	0.0643	0.0188	0.0634	0.0047	0.0368	0.0015	-0.0049	0.0387	0.0395	-0.0224	-0.0356	-0.0277	0.0398	-0.0291	0.0643	0.0170	0.0361	-0.0039	-0.0181	-0.0083	-0.0030	-0.0241	0.0225	-0.0190	-0.0051	-0.0033	0.0445	0.0683	0.0932	0.0529	-0.0012	0.0152	0.0433	0.0101	-0.0015	0.0000	-0.0169	-0.0260
37	0.0266	-0.0012	0.0532	0.0130	0.0648	-0.0261	0.0072	0.0150	0.0236	-0.0177	-0.0122	-0.0254	0.0250	0.0021	0.0498	-0.0155	0.0085	-0.0341	-0.0328	-0.0201	0.0157	-0.0241	0.0340	-0.0199	-0.0106	-0.0165	0.0438	0.0596	0.0980	0.0541	-0.0130	-0.0252	0.0582	0.0088	0.0200	0.0169	0.0000	-0.0020
38	0.0660	0.0232	0.0729	0.0355	0.0387	0.0154	0.0044	0.0569	0.0488	-0.0079	-0.0185	-0.0198	0.0469	0.0083	0.0735	0.0246	0.0074	0.0159	-0.0174	-0.0203	0.0367	-0.0109	0.0561	-0.0096	-0.0077	-0.0034	0.0591	0.0815	0.0850	0.0746	0.0020	0.0232	0.0487	0.0205	0.0168	0.0260	0.0020	0.0000

Table B.4: Magnitudes of the differences between δ_S of each dataset.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	0.0000	0.0355	0.0055	0.0263	0.0249	0.0471	0.0643	0.0054	0.0291	0.0254	0.0681	0.0769	0.0382	0.0335	0.0182	0.0214	0.0339	0.0733	0.0584	0.0631	0.0352	0.0820	0.0208	0.0646	0.0453	0.0343	0.0296	0.0232	0.0769	0.0411	0.0357	0.0530	0.0216	0.0378	0.0479	0.0643	0.0266	0.0660
2	0.0355	0.0000	0.0847	0.0134	0.0416	0.0069	0.0144	0.0629	0.0105	0.0007	0.0284	0.0231	0.0163	0.0006	0.1160	0.0106	0.0141	0.0070	0.0250	0.0134	0.0418	0.0200	0.0226	0.0243	0.0059	0.0033	0.0795	0.0631	0.0678	0.0318	0.0027	0.0106	0.0484	0.0025	0.0209	0.0188	0.0012	0.0232
3	0.0055	0.0847	0.0000	0.0332	0.0027	0.0905	0.0895	0.0233	0.0407	0.0595	0.0715	0.0679	0.0399	0.0367	0.0086	0.0156	0.0548	0.0977	0.1031	0.1003	0.0350	0.0942	0.0294	0.0549	0.0794	0.0476	0.0066	0.0278	0.0618	0.0494	0.0601	0.0797	0.0162	0.0235	0.0754	0.0634	0.0532	0.0729
4	0.0263	0.0134	0.0332	0.0000	0.0652	0.0032	0.0049	0.0111	0.0322	0.0459	0.0145	0.0383	0.0499	0.0137	0.0236	0.0106	0.0147	0.0095	0.0297	0.0418	0.0296	0.0356	0.0200	0.0359	0.0044	0.0214	0.0128	0.0445	0.1025	0.0695	0.0103	0.0192	0.0159	0.0153	0.0260	0.0047	0.0130	0.0355
5	0.0249	0.0416	0.0027	0.0652	0.0000	0.0528	0.0890	0.0298	0.0211	0.0699	0.0901	0.0798	0.0173	0.0041	0.0181	0.0402	0.0460	0.0454	0.0847	0.0739	0.0315	0.0456	0.0141	0.1193	0.0609	0.0713	0.0177	0.0406	0.0328	0.0333	0.0824	0.0405	0.0052	0.0561	0.0482	0.0368	0.0648	0.0387
6	0.0471	0.0069	0.0905	0.0032	0.0528	0.0000	0.0065	0.0353	0.0343	0.0132	0.0190	0.0255	0.0504	0.0003	0.0767	0.0236	0.0204	0.0114	0.0224	0.0053	0.0013	0.0019	0.0108	0.0206	0.0083	0.0071	0.0481	0.0565	0.0808	0.0340	0.0084	0.0147	0.0700	0.0053	0.0153	0.0015	0.0261	0.0154
7	0.0643	0.0144	0.0895	0.0049	0.0890	0.0065	0.0000	0.0491	0.0534	0.0231	0.0111	0.0222	0.0608	0.0137	0.0860	0.0062	0.0361	0.0084	0.0214	0.0051	0.0068	0.0048	0.0064	0.0228	0.0101	0.0017	0.0651	0.0607	0.1311	0.0581	0.0236	0.0127	0.0832	0.0039	0.0224	0.0049	0.0072	0.0044
8	0.0054	0.0629	0.0233	0.0111	0.0298	0.0353	0.0491	0.0000	0.0066	0.0526	0.0533	0.0615	0.0065	0.0495	0.0238	0.0137	0.0311	0.0282	0.0567	0.0443	0.0036	0.0670	0.0096	0.0439	0.0385	0.0344	0.0539	0.0440	0.0448	0.0180	0.0271	0.0271	0.0058	0.0258	0.0292	0.0387	0.0150	0.0569
9	0.0291	0.0105	0.0407	0.0322	0.0211	0.0343	0.0534	0.0066	0.0000	0.0329	0.0128	0.0390	0.0034	0.0196	0.0323	0.0161	0.0179	0.0325	0.0425	0.0336	0.0076	0.0790	0.0037	0.0324	0.0351	0.0322	0.0316	0.0546	0.0853	0.0579	0.0359	0.0523	0.0065	0.0256	0.0304	0.0395	0.0236	0.0488
10	0.0254	0.0007	0.0595	0.0459	0.0699	0.0132	0.0231	0.0526	0.0329	0.0000	0.0174	0.0183	0.0282	0.0106	0.0317	0.0523	0.0123	0.0118	0.0012	0.0038	0.0216	0.0025	0.0624	0.0193	0.0308	0.0147	0.1249	0.0725	0.0951	0.0871	0.0205	0.0189	0.0285	0.0202	0.0241	0.0224	0.0177	0.0079
11	0.0681	0.0284	0.0715	0.0145	0.0901	0.0190	0.0111	0.0533	0.0128	0.0174	0.0000	0.0107	0.0218	0.0281	0.0881	0.0276	0.0373	0.0155	0.0036	0.0204	0.0217	0.0225	0.0268	0.0075	0.0132	0.0093	0.0790	0.0779	0.1434	0.0647	0.0088	0.0347	0.0392	0.0281	0.0204	0.0356	0.0122	0.0185
12	0.0769	0.0231	0.0679	0.0383	0.0798	0.0255	0.0222	0.0615	0.0390	0.0183	0.0107	0.0000	0.0283	0.0062	0.0917	0.0557	0.0505	0.0169	0.0071	0.0186	0.0287	0.0194	0.0847	0.0598	0.0075	0.0235	0.0960	0.1036	0.0806	0.0815	0.0181	0.0362	0.0283	0.0417	0.0363	0.0277	0.0254	0.0198
13	0.0382	0.0163	0.0399	0.0499	0.0173	0.0504	0.0608	0.0065	0.0034	0.0123	0.0218	0.0283	0.0000	0.0156	0.0445	0.0225	0.0211	0.0315	0.0689	0.0363	0.0945	0.0847	0.0005	0.0251	0.0478	0.0356	0.0326	0.0576	0.0800	0.0739	0.0341	0.0645	0.0013	0.0337	0.0296	0.0398	0.0250	0.0469
14	0.0335	0.0006	0.0367	0.0137	0.0041	0.0003	0.0137	0.0495	0.0196	0.0106	0.0281	0.0062	0.0156	0.0000	0.0891	0.0199	0.0294	0.0011	0.0342	0.0258	0.0116	0.0140	0.0281	0.0100	0.0024	0.0217	0.0473	0.0386	0.1179	0.0349	0.0222	0.0058	0.0012	0.0133	0.0150	0.0291	0.0021	0.0083
15	0.0182	0.1160	0.0086	0.0236	0.0181	0.0767	0.0860	0.0238	0.0329	0.0329	0.0317	0.0881	0.0917	0.0445	0.0891	0.0000	0.0653	0.0771	0.0829	0.0599	0.0409	0.0779	0.0355	0.0815	0.0696	0.0467	0.0372	0.0032	0.0223	0.0165	0.0607	0.0602	0.0327	0.0244	0.0589	0.0643	0.0498	0.0735
16	0.0214	0.0106	0.0156	0.0106	0.0402	0.0236	0.0062	0.0137	0.0311	0.0179	0.0523	0.0557	0.0225	0.0199	0.0653	0.0000	0.0053	0.0058	0.0249	0.0125	0.0292	0.0268	0.0089	0.0164	0.0129	0.0071	0.0474	0.0397	0.0825	0.0350	0.0009	0.0111	0.0623	0.0158	0.0103	0.0170	0.0155	0.0246
17	0.0339	0.0141	0.0548	0.0147	0.0460	0.0204	0.0361	0.0311	0.0179	0.0123	0.0373	0.0505	0.0211	0.0294	0.0771	0.0953	0.0000	0.0144	0.0231	0.0182	0.0974	0.0177	0.0131	0.0388	0.0109	0.0233	0.0283	0.0599	0.0794	0.0641	0.0150	0.0062	0.0352	0.0096	0.0089	0.0361	0.0085	0.0074
18	0.0733	0.0070	0.0977	0.0095	0.0454	0.0114	0.0084	0.0282	0.0325	0.0118	0.0155	0.0169	0.0315	0.0011	0.0736	0.0058	0.0144	0.0000	0.0141	0.0036	0.0221	0.0013	0.0089	0.0183	0.0061	0.0033	0.0540	0.0630	0.0857	0.0505	0.0092	0.0209	0.0613	0.0101	0.0173	0.0039	0.0341	0.0159
19	0.0584	0.0250	0.1031	0.0297	0.0847	0.0224	0.0214	0.0567	0.0425	0.0012	0.0036	0.0071	0.0689	0.0342	0.0829	0.0249	0.0231	0.0141	0.0000	0.0059	0.0599	0.0062	0.0373	0.0073	0.0194	0.0013	0.0563	0.0748	0.1372	0.0663	0.0283	0.0232	0.0844	0.0219	0.0398	0.0181	0.0328	0.0174
20	0.0631	0.0134	0.1003	0.0418	0.0739	0.0053	0.0051	0.0443	0.0336	0.0204	0.0186	0.0363	0.0258	0.0799	0.0125	0.0182	0.0036	0.0059	0.0000	0.0490	0.0000	0.0490	0.0097	0.0462	0.0131	0.0161	0.0614	0.0765	0.0808	0.0808	0.0137	0.0000	0.0699	0.0001	0.0228	0.0083	0.0201	0.0203
21	0.0352	0.0418	0.0350	0.0296	0.0315	0.0013	0.0068	0.0036	0.0076	0.0216	0.0217	0.0287	0.0045	0.0116	0.0409	0.0292	0.0074	0.0221	0.0599	0.0490	0.0000	0.0545	0.0142	0.0187	0.0276	0.0140	0.0293	0.0417	0.1007	0.0646	0.0133	0.0140	0.0220	0.0228	0.0028	0.0030	0.0157	0.0367
22	0.0820	0.0200	0.0942	0.0356	0.0456	0.0019	0.0048	0.0670	0.0790	0.0025	0.0225	0.0194	0.0847	0.0140	0.0779	0.0268	0.0177	0.0013	0.0062	0.0097	0.0545	0.0000	0.0027	0.0104	0.0108	0.0021	0.0528	0.0573	0.0968	0.0694	0.0178	0.0143	0.0573	0.0171	0.0207	0.0241	0.0241	0.0109
23	0.0208	0.0226	0.0294	0.0200	0.0141	0.0108	0.0064	0.0006	0.0037	0.0624	0.0268	0.0598	0.0005	0.0281	0.0355	0.0089	0.0131	0.0089	0.0373	0.0462	0.0142	0.0027	0.0000	0.0546	0.0374	0.0468	0.0314	0.0229	0.0696	0.0227	0.0382	0.0042	0.0061	0.0119	0.0410	0.0225	0.0340	0.0561
24	0.0646	0.0243	0.0549	0.0359	0.1193	0.0206	0.0228	0.0439	0.0324	0.0193	0.0075	0.0075	0.0251	0.0100	0.0815	0.0164	0.0388	0.0183	0.0073	0.0131	0.0187	0.0104	0.0546	0.0000	0.0216	0.0155	0.0885	0.0716	0.0728	0.0805	0.0108	0.0358	0.0254	0.0228	0.0293	0.0190	0.0199	0.0096
25	0.0453	0.0059	0.0794	0.0044	0.0609	0.0083	0.0101	0.0385	0.0351	0.0308	0.0132	0.0274	0.0478	0.0024	0.0696	0.0129	0.0109	0.0061	0.0194	0.0161	0.0276	0.0108	0.0374	0.0216	0.0000	0.0161	0.0578	0.0568	0.1251	0.0480	0.0043	0.0137	0.0509	0.0080	0.0173	0.0051	0.0106	0.0077
26	0.0343	0.0033	0.0476	0.0214	0.0713	0.0071	0.0017	0.0344	0.0322	0.0147	0.0093	0.0235	0.0356	0.0217	0.0467	0.0071	0.0233	0.0033	0.0013	0.0040	0.0140	0.0021	0.0468	0.0155	0.0161	0.0000	0.0614	0.0568	0.0807	0.0772	0.0033	0.0177	0.0630	0.0150	0.0066	0.0033	0.0165	0.0034
27	0.0296	0.0795	0.0066	0.0128	0.0177	0.0481	0.0651	0.0539	0.0316	0.1249	0.0790	0.0960	0.0326	0.0473	0.0372	0.0474	0.0283	0.0540	0.0563	0.0614	0.0293	0.0528	0.0314	0.0885	0.0578	0.0614	0.0000	0.0251	0.0251	0.0019	0.0566	0.0383	0.0038	0.0466	0.0397	0.0445	0.0438	0.0591
28	0.0232	0.0631	0.0278	0.0445	0.0406	0.0565	0.0907	0.0440	0.0853	0.0951	0.1434	0.0806	0.0800	0.1179	0.0223	0.0825	0.0794	0.0857	0.1372	0.0808	0.1007	0.0968	0.0696	0.0728	0.1251	0.0807	0.0251	0.0213	0.0000	0.0348	0.1096	0.0710	0.0377	0.0849	0.1096	0.0932	0.0980	0.0850
29	0.0769	0.0678	0.0618	0.1025	0.0328	0.0808	0.1311	0.0448	0.0853	0.0951	0.1434	0.0806	0.0800	0.1179	0.0223	0.0825	0.0794	0.0857	0.1372	0.0808	0.1007	0.0968	0.0696	0.0728	0.1251	0.0807	0.0251	0.0213	0.0000	0.0348	0.1096	0.0710	0.0377	0.0849	0.1096	0.0932	0.0980	0.0850
30	0.0411	0.0318	0.0494	0.0695	0.0333	0.0340	0.0581	0.0180	0.0579	0.0871	0.0647	0.0815	0.0739	0.0349	0.0165	0.0350	0.0641	0.0505	0.0663	0.0808	0.0646	0.0694	0.0227	0.0805	0.0480	0.0772	0.0019	0.0162	0.0348	0.0000	0.0836	0.0382	0.0423	0.0330	0.0531	0.0529	0.0541	0.0746
31	0.0357	0.0027	0.0601	0.0103	0.0824	0.0084	0.0236	0.0271	0.0359	0.0205	0.0088	0.0181	0.0341	0.0222	0.0607	0.0909	0.0150	0.0092	0.0283	0.0137	0.0133	0.0178	0.0382	0.0108	0.0043	0.0033	0.0566	0.0796	0.1096	0.0836	0.0000	0.0977	0.0387	0.0046	0.0028	0.0012	0.0130	0.0020
32	0.0530	0.0106	0.0797	0.0192	0.0405	0.0147	0.0127	0.0271	0.0523	0.0189	0.0347	0.0362	0.0645	0.0058	0.0602	0.0111	0.0062	0.0209	0.0232	0.0000	0.0140	0.0143	0.0042	0.0358	0.0137	0.0177	0.0383	0.0513	0.0710	0.0382	0.0077	0.0000	0.0427	0.0127	0.0171	0.0152	0.0252	0.0232
33	0.0216	0.0484	0.0162	0.0159	0.0952	0.0700	0.0832	0.0058	0.0065	0.0285	0.0392	0.0283	0.0013	0.0012	0.0327	0.0623	0.0352	0.0613	0.0844	0.0699	0.0220	0.0573	0.0061	0.0254	0.0509	0.0630	0.0038	0.0260	0.0377	0.0423	0.0387	0.0427	0.0000	0.0702	0.0379	0.0433	0.0582	0.0487
34	0.0378	0.0025	0.0235	0.0153	0.0561	0.0053	0.0039	0.0258	0.0256	0.0202	0.0281	0.0417	0.0337	0.0133	0.0244	0.0158	0.0096	0.0101	0.0219	0.0001	0.0228	0.0171	0.0119	0.0228	0.0080	0.0150	0.0466	0.0541	0.0849	0.0330	0.0046	0.0127	0.0702	0.0000	0.0100	0.0101	0.0088	0.0205
35	0.0479	0.0209	0.0754	0.0260	0.0482	0.0153	0.0224	0.0292	0.0304	0.0241	0.0204	0.0363	0.0296	0.0150	0.0589	0.0103	0.0089	0.0173	0.0398	0.0228	0.0207	0.0410	0.0293	0.0173	0.0066	0.0397	0.0531	0.1096	0.0531	0.0028	0.0171	0.0379	0.0100	0.0000	0.0015	0.0200	0.0168
36	0.0643	0.0188	0.0634	0.0047	0.0368	0.0015	0.0049	0.0387	0.0395	0.0224	0.0356	0.0277	0.0398	0.0291	0.0643	0.0170	0.0361	0.0039	0.0181	0.0083	0.0030	0.0241	0.0225	0.0190	0.0051	0.0033	0.0445	0.0683	0.0932	0.0529	0.0012	0.0152	0.0433	0.0101	0.0015	0.0000	0.0169	0.0260
37	0.0266	0.0012	0.0532	0.0130	0.0648	0.0261	0.0072	0.0150	0.0236	0.0177	0.0122	0.0254	0.0250	0.0021	0.0498	0.0155	0.0085	0.0341	0.0328	0.0201	0.0157	0.0241	0.0340	0.0199	0.0106	0.0165	0.0438	0.0596	0.0980	0.0541	0.0130	0.0252	0.0582	0.0088	0.0200	0.0169	0.0000	0.0020
38	0.0660	0.0232	0.0729	0.0355	0.0387	0.0154	0.0044	0.0569	0.0488	0.0979	0.0185	0.0198	0.0469	0.0083	0.0735	0.0246	0.0074	0.0159	0.0174	0.0203	0.0367	0.0109	0.0561	0.0096	0.0077	0.0034	0.0591	0.0815	0.0850	0.0746	0.0020	0.0232	0.0487	0.0205	0.0168	0.0260	0.0020	0.0000

Table B.5: Differences between $\delta_{|F|}$ of each dataset.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	0.0000	0.0519	0.0102	0.0280	0.0249	0.0471	0.0643	0.0176	0.0291	0.0254	0.0681	0.0769	0.0382	0.0629	0.0182	0.0214	0.0339	0.0733	0.0584	0.0631	0.0352	0.0820	0.0219	0.0646	0.0453	0.0343	0.0647	0.1188	0.1384	0.1010	0.0357	0.0530	0.0216	0.0378	0.0479	0.0643	0.0266	0.0660
2	0.0519	0.0000	0.0847	0.0134	0.0416	0.0157	0.0180	0.0629	0.0351	0.0193	0.0284	0.0265	0.0451	0.0584	0.1160	0.0209	0.0743	0.0462	0.0378	0.0355	0.0418	0.0361	0.0458	0.0243	0.0435	0.0234	0.1097	0.1185	0.1060	0.0764	0.0456	0.0560	0.0484	0.0313	0.0209	0.0414	0.0329	0.0541
3	0.0102	0.0847	0.0000	0.0734	0.0227	0.0905	0.0895	0.0233	0.0407	0.0595	0.0715	0.0679	0.0399	0.0480	0.0132	0.0417	0.0548	0.0977	0.1031	0.1003	0.0350	0.0942	0.0294	0.0549	0.0794	0.0476	0.0556	0.0278	0.1137	0.0494	0.0601	0.0797	0.0208	0.0440	0.0754	0.0634	0.0532	0.0729
4	0.0280	0.0134	0.0734	0.0000	0.0652	0.0061	0.0095	0.0256	0.0559	0.0459	0.0145	0.0383	0.0710	0.0410	0.0429	0.0230	0.0419	0.0295	0.0297	0.0418	0.0296	0.0356	0.0200	0.0359	0.0285	0.0214	0.0804	0.0647	0.1376	0.0695	0.0847	0.0384	0.0663	0.0369	0.0260	0.0264	0.0130	0.0355
5	0.0471	0.0157	0.0905	0.0061	0.0528	0.0000	0.0105	0.0353	0.0343	0.0132	0.0190	0.0255	0.0607	0.0124	0.0767	0.0236	0.0445	0.0168	0.0290	0.0181	0.0031	0.0161	0.0144	0.0206	0.0225	0.0071	0.0570	0.0932	0.1226	0.0549	0.0498	0.0438	0.0700	0.0144	0.0234	0.0140	0.0361	0.0154
6	0.0249	0.0416	0.0227	0.0652	0.0000	0.0528	0.0890	0.0494	0.0249	0.0699	0.0901	0.0798	0.0322	0.0377	0.0219	0.0402	0.0460	0.0454	0.0847	0.0739	0.0315	0.0456	0.0157	0.1193	0.0609	0.0713	0.1043	0.0758	0.1765	0.0587	0.0824	0.0405	0.0852	0.0561	0.0482	0.0368	0.0648	0.0387
7	0.0643	0.0180	0.0895	0.0095	0.0890	0.0105	0.0000	0.0491	0.0534	0.0231	0.0111	0.0222	0.0608	0.0137	0.0860	0.0062	0.0364	0.0084	0.0214	0.0163	0.0068	0.0048	0.0076	0.0228	0.0101	0.0250	0.0893	0.0934	0.1791	0.0787	0.0277	0.0361	0.0832	0.0039	0.0224	0.0122	0.0073	0.0044
8	0.0176	0.0629	0.0233	0.0256	0.0494	0.0353	0.0491	0.0000	0.0114	0.0526	0.0533	0.0615	0.0219	0.0495	0.0238	0.0137	0.0311	0.0282	0.0567	0.0443	0.0036	0.0670	0.0122	0.0498	0.0385	0.0344	0.0907	0.0845	0.0767	0.0480	0.0309	0.0431	0.0132	0.0258	0.0292	0.0387	0.0150	0.0569
9	0.0291	0.0351	0.0407	0.0559	0.0249	0.0343	0.0534	0.0114	0.0000	0.0329	0.0128	0.0390	0.0180	0.0468	0.0323	0.0271	0.0179	0.0325	0.0425	0.0336	0.0076	0.0790	0.0098	0.0324	0.0351	0.0322	0.0468	0.0830	0.1338	0.0794	0.0359	0.0523	0.0181	0.0400	0.0304	0.0395	0.0236	0.0488
10	0.0254	0.0193	0.0595	0.0459	0.0699	0.0132	0.0231	0.0526	0.0329	0.0000	0.0174	0.0183	0.0294	0.0233	0.0317	0.0523	0.0123	0.0169	0.0150	0.0044	0.0239	0.0200	0.0624	0.0193	0.0308	0.0147	0.1249	0.0806	0.0955	0.0871	0.0331	0.0272	0.0285	0.0258	0.0241	0.0224	0.0230	0.0223
11	0.0681	0.0284	0.0715	0.0145	0.0901	0.0190	0.0111	0.0533	0.0128	0.0174	0.0000	0.0128	0.0218	0.0281	0.0881	0.0276	0.0373	0.0155	0.0109	0.0204	0.0217	0.0225	0.0268	0.0075	0.0165	0.0093	0.0837	0.0888	0.1659	0.0724	0.0262	0.0389	0.0392	0.0281	0.0204	0.0356	0.0122	0.0185
12	0.0769	0.0265	0.0679	0.0383	0.0798	0.0255	0.0222	0.0615	0.0390	0.0183	0.0128	0.0000	0.0283	0.0662	0.0917	0.0557	0.0505	0.0169	0.0084	0.0186	0.0287	0.0194	0.0598	0.0107	0.0274	0.0235	0.0960	0.1054	0.0830	0.0815	0.0204	0.0362	0.0283	0.0417	0.0363	0.0277	0.0254	0.0198
13	0.0382	0.0451	0.0399	0.0710	0.0322	0.0607	0.0608	0.0219	0.0180	0.0294	0.0218	0.0283	0.0000	0.0304	0.0445	0.0461	0.0211	0.0489	0.0689	0.0388	0.0260	0.0847	0.0193	0.0251	0.0478	0.0377	0.0387	0.0758	0.0949	0.0810	0.0341	0.0645	0.0316	0.0669	0.0427	0.0398	0.0250	0.0469
14	0.0629	0.0584	0.0480	0.0410	0.0377	0.0124	0.0137	0.0495	0.0468	0.0233	0.0281	0.0062	0.0304	0.0000	0.0891	0.0248	0.0294	0.0043	0.0342	0.0258	0.0239	0.0246	0.0341	0.0100	0.0269	0.0217	0.0669	0.0812	0.1347	0.0573	0.0222	0.0382	0.0388	0.0133	0.0260	0.0291	0.0115	0.0163
15	0.0182	0.1160	0.0132	0.0429	0.0219	0.0767	0.0860	0.0238	0.0323	0.0317	0.0881	0.0917	0.0445	0.0891	0.0000	0.0653	0.0771	0.0716	0.0829	0.0799	0.0409	0.0779	0.0355	0.0815	0.0696	0.0467	0.0472	0.0941	0.0914	0.0435	0.0607	0.0602	0.0327	0.0257	0.0589	0.0643	0.0498	0.0735
16	0.0214	0.0209	0.0417	0.0230	0.0402	0.0236	0.0062	0.0137	0.0271	0.0523	0.0276	0.0557	0.0461	0.0248	0.0653	0.0000	0.0207	0.0103	0.0249	0.0125	0.0292	0.0268	0.0089	0.0164	0.0129	0.0071	0.0474	0.0581	0.0986	0.0422	0.0238	0.0273	0.0823	0.0173	0.0391	0.0170	0.0155	0.0246
17	0.0339	0.0743	0.0548	0.0419	0.0460	0.0445	0.0364	0.0311	0.0179	0.0123	0.0373	0.0505	0.0211	0.0294	0.0771	0.0207	0.0000	0.0244	0.0231	0.0182	0.0159	0.0394	0.0131	0.0388	0.0219	0.0088	0.0376	0.0811	0.1046	0.0759	0.0150	0.0962	0.0380	0.0319	0.0273	0.0361	0.0085	0.0223
18	0.0733	0.0462	0.0977	0.0295	0.0454	0.0168	0.0084	0.0282	0.0325	0.0169	0.0155	0.0169	0.0489	0.0043	0.0736	0.0103	0.0244	0.0000	0.0141	0.0130	0.0430	0.0013	0.0089	0.0183	0.0064	0.0088	0.0540	0.0783	0.1058	0.0564	0.0267	0.0322	0.0613	0.0101	0.0319	0.0039	0.0341	0.0159
19	0.0584	0.0378	0.1031	0.0297	0.0847	0.0290	0.0214	0.0567	0.0425	0.0150	0.0109	0.0084	0.0689	0.0342	0.0829	0.0249	0.0231	0.0141	0.0000	0.0107	0.0599	0.0149	0.0404	0.0082	0.0194	0.0118	0.0563	0.0748	0.1426	0.0663	0.0289	0.0232	0.0844	0.0219	0.0398	0.0181	0.0328	0.0174
20	0.0631	0.0355	0.1003	0.0418	0.0739	0.0181	0.0163	0.0443	0.0336	0.0044	0.0204	0.0186	0.0388	0.0258	0.0799	0.0125	0.0182	0.0130	0.0107	0.0000	0.0490	0.0000	0.0545	0.0195	0.0462	0.0141	0.0263	0.0040	0.0618	0.0835	0.0355	0.0158	0.0699	0.0053	0.0228	0.0083	0.0201	0.0203
21	0.0352	0.0418	0.0350	0.0296	0.0315	0.0031	0.0068	0.0036	0.0076	0.0239	0.0217	0.0287	0.0260	0.0239	0.0409	0.0292	0.0159	0.0430	0.0599	0.0490	0.0000	0.0545	0.0195	0.0462	0.0000	0.0304	0.0588	0.0980	0.1220	0.0672	0.0133	0.0526	0.0220	0.0228	0.0080	0.0211	0.0157	0.0367
22	0.0820	0.0361	0.0942	0.0356	0.0456	0.0161	0.0048	0.0670	0.0790	0.0200	0.0225	0.0194	0.0847	0.0246	0.0779	0.0268	0.0394	0.0013	0.0149	0.0097	0.0545	0.0000	0.0062	0.0104	0.0108	0.0087	0.0528	0.0761	0.1018	0.0741	0.0341	0.0311	0.0573	0.0171	0.0344	0.0241	0.0241	0.0109
23	0.0219	0.0458	0.0294	0.0200	0.0157	0.0144	0.0076	0.0122	0.0098	0.0624	0.0268	0.0598	0.0193	0.0341	0.0355	0.0089	0.0131	0.0089	0.0404	0.0462	0.0195	0.0062	0.0000	0.0546	0.0374	0.0468	0.0415	0.0483	0.1273	0.0395	0.0382	0.0327	0.0889	0.0119	0.0410	0.0225	0.0340	0.0561
24	0.0646	0.0243	0.0549	0.0359	0.1193	0.0206	0.0228	0.0439	0.0324	0.0193	0.0075	0.0107	0.0251	0.0100	0.0815	0.0164	0.0388	0.0183	0.0082	0.0141	0.0187	0.0104	0.0546	0.0000	0.0216	0.0155	0.0885	0.0727	0.0897	0.0805	0.0255	0.0358	0.0254	0.0228	0.0293	0.0190	0.0199	0.0096
25	0.0453	0.0435	0.0794	0.0285	0.0609	0.0225	0.0101	0.0385	0.0351	0.0308	0.0165	0.0274	0.0478	0.0269	0.0696	0.0129	0.0109	0.0064	0.0194	0.0263	0.0276	0.0108	0.0374	0.0216	0.0000	0.0216	0.0578	0.0653	0.1371	0.0506	0.0185	0.0137	0.0509	0.0080	0.0173	0.0096	0.0106	0.0077
26	0.0343	0.0234	0.0476	0.0214	0.0713	0.0071	0.0250	0.0344	0.0322	0.0147	0.0093	0.0235	0.0377	0.0217	0.0467	0.0071	0.0233	0.0088	0.0118	0.0040	0.0304	0.0087	0.0468	0.0155	0.0216	0.0000	0.0616	0.0738	0.1048	0.0772	0.0369	0.0302	0.0630	0.0150	0.0083	0.0118	0.0208	0.0158
27	0.0647	0.1097	0.0556	0.0804	0.1043	0.0570	0.0893	0.0907	0.0468	0.1249	0.0837	0.0960	0.0387	0.0669	0.0472	0.0376	0.0540	0.0563	0.0618	0.0588	0.0528	0.0415	0.0885	0.0578	0.0616	0.0000	0.0384	0.0444	0.0082	0.0566	0.0383	0.0938	0.0466	0.0788	0.0510	0.0513	0.0683
28	0.1188	0.1185	0.0278	0.0647	0.0758	0.0932	0.0934	0.0845	0.0830	0.0806	0.0888	0.1054	0.0758	0.0812	0.0941	0.0581	0.0811	0.0783	0.0748	0.0835	0.0980	0.0761	0.0483	0.0727	0.0653	0.0738	0.0384	0.0000	0.0213	0.0331	0.0796	0.0513	0.1284	0.0799	0.1033	0.0990	0.0743	0.1059
29	0.1384	0.1090	0.1137	0.1376	0.1765	0.1226	0.1791	0.0767	0.1338	0.0955	0.1659	0.0830	0.0949	0.1347	0.0914	0.0986	0.1046	0.1058	0.1426	0.1001	0.1220	0.1018	0.1273	0.0897	0.1371	0.1048	0.0444	0.0213	0.0000	0.0622	0.1096	0.0882	0.1600	0.1160	0.1502	0.1020	0.1262	0.0962
30	0.1010	0.0764	0.0494	0.0695	0.0887	0.0549	0.0787	0.0480	0.0794	0.0871	0.0724	0.0815	0.0810	0.0573	0.0435	0.0422	0.0759	0.0564	0.0663	0.0808	0.0672	0.0741	0.0395	0.0805	0.0506	0.0772	0.0082	0.0331	0.0622	0.0000	0.0836	0.0382	0.1044	0.0436	0.0887	0.0613	0.0552	0.0807
31	0.0357	0.0456	0.0601	0.0847	0.0824	0.0498	0.0277	0.0309	0.0359	0.0331	0.0262	0.0204	0.0341	0.0222	0.0607	0.0238	0.0150	0.0267	0.0289	0.0355	0.0133	0.0341	0.0382	0.0255	0.0185	0.0369	0.0566	0.0796	0.1096	0.0836	0.0000	0.0162	0.0387	0.0290	0.0513	0.0261	0.0203	0.0257
32	0.0530	0.0560	0.0797	0.0384	0.0405	0.0438	0.0361	0.0431	0.0523	0.0272	0.0389	0.0362	0.0645	0.0382	0.0602	0.0273	0.0062	0.0322	0.0232	0.0158	0.0526	0.0311	0.0327	0.0358	0.0185	0.0302	0.0383	0.0513	0.0882	0.0382	0.0162	0.0000	0.0427	0.0319	0.0576	0.0283	0.0252	0.0357
33	0.0216	0.0484	0.0208	0.0663	0.0952	0.0700	0.0832	0.0132	0.0181	0.0285	0.0392	0.0283	0.0316	0.0388	0.0327	0.0623	0.0380	0.0613	0.0844	0.0699	0.0220	0.0573	0.0089	0.0254	0.0509	0.0630	0.0938	0.1284	0.1600	0.1044	0.0387	0.0427	0.0000	0.0702	0.0379	0.0433	0.0582	0.0487
34	0.0378	0.0313	0.0440	0.0309	0.0561	0.0144	0.0039	0.0258	0.0400	0.0258	0.0281	0.0417	0.0069	0.0133	0.0257	0.0173	0.0319	0.0101	0.0219	0.0053	0.0228	0.0171	0.0119	0.0228	0.0080	0.0150	0.0466	0.0799	0.1160	0.0436	0.0290	0.0319	0.0702	0.0000	0.0310	0.0101	0.0088	0.0205
35	0.0479	0.0209	0.0754	0.0260	0.0482	0.0234	0.0224	0.0292	0.0304	0.0241	0.0204	0.0363	0.0427	0.0260	0.0589	0.0391	0.0273	0.0319	0.0398	0.0228	0.0690	0.0344	0.0410	0.0293	0.0173	0.0083	0.0788	0.1033	0.1592	0.0887	0.0513	0.0576	0.0379	0.0310	0.0000	0.0284	0.0200	0.0388
36	0.0643	0.0414	0.0634	0.0264	0.0368	0.0140	0.0122	0.0387	0.0395	0.0224	0.0356	0.0277	0.0398	0.0291	0.0643	0.0170	0.0361	0.0039	0.0181	0.0083	0.0211	0.0241	0.0225	0.0190	0.0096	0.0118	0.0510	0.0930	0.1020	0.0613	0.0261	0.0283	0.0433	0.0101	0.0284	0.0000	0.0169	0.0260
37	0.0266	0.0329	0.0532	0.0130	0.0648	0.0361	0.0073	0.0150	0.0236	0.0230	0.0122	0.0254	0.0250	0.0115	0.0498	0.0155	0.0085	0.0341	0.0328	0.0201	0.0157	0.0241	0.0340	0.0199	0.0106	0.0208	0.0513	0.0743	0.1262	0.0552	0.0203	0.0252	0.0582	0.0088	0.0200	0.0169	0.0000	0.0140
38	0.0660	0.0541	0.0729	0.0355	0.0387	0.0154	0.0044	0.0569	0.0488	0.0223	0.0185	0.0198	0.0469	0.0163	0.0735	0.0246	0.0222	0.0159	0.0174	0.0203	0.0367	0.0109	0.0561	0.0096	0.0077	0.0158	0.0683	0.1059	0.0962	0.0807	0.0257	0.0357	0.0487	0.0205	0.0388	0.0260	0.0140	0.0000

Table B.6: Procrustes distance between each dataset.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
1	0.0000	0.9991	0.9986	0.9981	0.9978	0.9990	0.9989	0.9973	0.9980	0.9977	0.9995	0.9979	0.9977	0.9995	0.9970	0.9971	0.9969	0.9967	0.9978	0.9998	0.9990	0.9997	0.9992	0.9987	0.9978	0.9990	0.9999	0.9991	0.9987	0.9993	0.9970	0.9991	0.9965	0.9985	0.9995	0.9996	0.9994	0.9998
2	0.9991	0.0000	0.9991	0.9988	0.9990	0.9974	0.9997	0.9978	0.9991	0.9967	0.9994	0.9988	0.9954	0.9981	0.9955	0.9974	0.9981	0.9994	0.9989	0.9990	0.9992	0.9999	0.9994	0.9997	0.9997	0.9997	0.9997	0.9978	0.9997	0.9990	0.9986	0.9996	0.9996	0.9999	0.9993	0.9969	0.9940	0.9942
3	0.9986	0.9991	0.0000	0.9988	0.9956	0.9969	0.9968	0.9999	0.9996	0.9987	0.9999	0.9997	0.9985	0.9974	0.9980	0.9994	0.9998	0.9990	0.9975	0.9972	0.9993	0.9992	0.9983	0.9980	0.9981	0.9990	0.9993	0.9985	0.9976	0.9991	0.9993	0.9994	0.9970	0.9993	0.9983	0.9996	0.9995	0.9990
4	0.9981	0.9988	0.9988	0.0000	0.9994	0.9996	0.9977	0.9983	0.9993	0.9992	0.9996	0.9993	0.9995	0.9992	0.9997	0.9986	0.9955	0.9987	0.9989	0.9977	0.9970	0.9992	0.9994	0.9992	0.9987	0.9967	0.9973	0.9957	0.9995	0.9980	0.9980	0.9972	0.9993	0.9997	0.9998	0.9959	0.9994	0.9995
5	0.9978	0.9990	0.9956	0.9994	0.0000	0.9999	0.9997	0.9990	0.9982	0.9983	0.9962	0.9995	0.9965	0.9934	0.9995	0.9971	0.9981	0.9976	0.9992	0.9965	0.9989	0.9977	0.9980	0.9985	0.9980	0.9989	0.9984	0.9974	0.9994	0.9995	0.9984	0.9996	0.9960	0.9984	0.9975	0.9984	0.9977	0.9988
6	0.9999	0.9974	0.9969	0.9996	0.9999	0.0000	0.9929	0.9998	0.9990	0.9985	0.9980	0.9992	0.9980	0.9978	0.9986	0.9977	0.9990	0.9983	0.9996	0.9988	0.9982	0.9984	0.9999	0.9998	0.9987	0.9991	0.9992	0.9990	0.9992	0.9980	0.9996	0.9987	0.9989	0.9994	0.9987	0.9997	0.9984	0.9996
7	0.9989	0.9997	0.9988	0.9977	0.9997	0.9929	0.0000	0.9985	0.9982	0.9977	0.9980	0.9964	0.9994	0.9989	0.9976	0.9974	0.9989	0.9987	0.9978	0.9993	0.9968	0.9992	0.9984	0.9998	0.9998	0.9995	0.9992	0.9966	0.9981	0.9980	0.9998	0.9986	0.9977	0.9996	0.9993	0.9996	0.9975	0.9993
8	0.9973	0.9978	0.9999	0.9983	0.9990	0.9998	0.9985	0.0000	0.9986	0.9974	0.9984	0.9987	0.9997	0.9998	0.9994	0.9996	0.9987	0.9997	0.9976	0.9991	0.9979	0.9992	0.9992	0.9996	0.9985	0.9989	0.9983	0.9983	0.9972	0.9995	0.9993	0.9988	0.9991	0.9989	0.9987	0.9986	0.9999	0.9987
9	0.9980	0.9978	0.9996	0.9993	0.9982	0.9990	0.9982	0.9986	0.0000	0.9997	0.9992	0.9987	0.9964	0.9998	0.9994	0.9996	0.9967	0.9973	0.9959	0.9983	0.9978	0.9991	0.9959	0.9996	0.9994	0.9991	0.9960	0.9984	0.9994	0.9991	0.9996	0.9995	0.9995	0.9997	0.9993	0.9996	0.9974	0.9989
10	0.9977	0.9991	0.9987	0.9992	0.9983	0.9985	0.9977	0.9974	0.9997	0.0000	0.9989	0.9961	0.9986	0.9993	0.9990	0.9997	0.9989	0.9990	0.9993	0.9982	0.9994	0.9980	0.9987	0.9987	0.9994	0.9996	0.9979	0.9989	0.9996	0.9989	0.9992	0.9992	0.9971	0.9979	1.0000	0.9982	0.9991	0.9990
11	0.9995	0.9967	0.9999	0.9996	0.9962	0.9980	0.9980	0.9984	0.9992	0.9989	0.0000	0.9994	0.9979	0.9992	0.9997	0.9991	0.9975	0.9982	0.9957	0.9993	0.9987	0.9988	0.9964	0.9984	0.9995	0.9983	0.9987	0.9994	0.9967	0.9975	0.9994	0.9994	0.9997	0.9987	0.9986	0.9996	0.9977	0.9982
12	0.9979	0.9994	0.9997	0.9993	0.9995	0.9992	0.9964	0.9987	0.9987	0.9961	0.9994	0.0000	0.9972	0.9965	0.9995	0.9988	0.9985	0.9996	0.9995	0.9992	0.9987	0.9997	0.9996	0.9990	0.9969	0.9997	0.9981	0.9994	0.9995	0.9959	0.9993	0.9997	0.9960	0.9987	0.9984	0.9980	0.9991	0.9996
13	0.9977	0.9988	0.9985	0.9995	0.9965	0.9980	0.9994	0.9997	0.9964	0.9986	0.9979	0.9972	0.0000	0.9996	0.9981	0.9990	0.9993	0.9986	0.9980	0.9981	0.9971	0.9994	0.9996	0.9997	0.9989	0.9993	0.9991	0.9994	0.9955	0.9999	0.9994	0.9979	0.9964	0.9989	0.9996	0.9994	0.9967	0.9994
14	0.9995	0.9954	0.9974	0.9992	0.9934	0.9978	0.9989	0.9989	0.9998	0.9993	0.9992	0.9965	0.9996	0.0000	0.9992	0.9993	0.9995	0.9977	0.9988	0.9980	0.9976	0.9998	0.9996	0.9989	0.9982	0.9983	0.9978	0.9989	0.9999	0.9999	0.9984	0.9985	0.9984	0.9970	0.9995	0.9996	0.9958	0.9982
15	0.9970	0.9981	0.9990	0.9997	0.9995	0.9986	0.9976	0.9979	0.9994	0.9990	0.9997	0.9995	0.9981	0.9992	0.0000	0.9993	0.9978	0.9995	0.9984	0.9981	0.9998	0.9991	0.9998	0.9978	0.9945	0.9983	0.9988	0.9990	0.9983	0.9988	0.9986	0.9982	0.9959	0.9979	0.9977	0.9998	0.9978	0.9992
16	0.9971	0.9955	0.9994	0.9986	0.9971	0.9977	0.9974	0.9996	0.9996	0.9997	0.9991	0.9988	0.9990	0.9993	0.9993	0.0000	0.9969	0.9984	0.9994	0.9991	0.9992	0.9986	0.9975	0.9991	0.9996	0.9971	0.9985	0.9992	0.9985	0.9985	0.9999	0.9991	0.9994	0.9995	0.9978	0.9995	0.9980	0.9992
17	0.9969	0.9974	0.9998	0.9955	0.9981	0.9990	0.9989	0.9987	0.9967	0.9989	0.9975	0.9985	0.9978	0.9969	0.9978	0.9969	0.0000	0.9997	0.9996	0.9993	0.9974	0.9949	0.9992	0.9985	0.9992	0.9996	0.9995	0.9964	0.9999	0.9971	0.9995	0.9994	0.9980	0.9996	0.9985	0.9992	0.9969	0.9992
18	0.9967	0.9981	0.9990	0.9987	0.9976	0.9983	0.9987	0.9997	0.9973	0.9990	0.9982	0.9987	0.9997	0.9997	0.9995	0.9984	0.9997	0.0000	0.9980	0.9996	0.9995	0.9988	0.9983	0.9990	0.9995	0.9991	0.9995	0.9991	0.9986	0.9990	0.9997	0.9992	0.9996	0.9979	0.9991	0.9974	0.9964	0.9983
19	0.9978	0.9994	0.9975	0.9989	0.9992	0.9996	0.9976	0.9959	0.9993	0.9957	0.9995	0.9992	0.9976	0.9988	0.9984	0.9994	0.9996	0.9980	0.0000	0.9988	0.9996	0.9993	0.9998	0.9990	0.9967	0.9974	0.9988	0.9977	0.9988	0.9989	0.9977	0.9988	0.9992	0.9988	0.9984	0.9982	0.9946	0.9985
20	0.9998	0.9989	0.9972	0.9977	0.9965	0.9988	0.9993	0.9991	0.9983	0.9982	0.9993	0.9992	0.9981	0.9992	0.9980	0.9991	0.9993	0.9996	0.9988	0.0000	0.9984	0.9996	0.9977	0.9986	0.9974	0.9986	0.9988	0.9970	0.9993	0.9992	0.9994	0.9994	0.9992	0.9992	0.9982	0.9988	0.9992	0.9976
21	0.9990	0.9990	0.9993	0.9970	0.9989	0.9982	0.9968	0.9979	0.9978	0.9994	0.9987	0.9987	0.9971	0.9976	0.9998	0.9992	0.9974	0.9995	0.9996	0.9984	0.0000	0.9986	0.9991	0.9976	0.9988	0.9986	0.9990	0.9996	0.9994	0.9998	0.9990	0.9990	0.9988	0.9989	0.9995	0.9989	0.9975	0.9993
22	0.9997	0.9992	0.9992	0.9992	0.9977	0.9984	0.9992	0.9992	0.9991	0.9988	0.9997	0.9994	0.9998	0.9991	0.9986	0.9949	0.9988	0.9993	0.9996	0.9986	0.0000	0.9943	0.9992	0.9994	0.9993	0.9994	0.9991	0.9990	0.9963	0.9994	0.9978	0.9980	0.9977	0.9961	0.9995	0.9964	0.9941
23	0.9992	0.9999	0.9983	0.9994	0.9980	0.9999	0.9984	0.9992	0.9960	0.9987	0.9964	0.9996	0.9996	0.9996	0.9975	0.9992	0.9983	0.9998	0.9977	0.9991	0.9943	0.0000	0.9994	0.9991	0.9973	0.9965	0.9996	0.9974	0.9991	0.9996	0.9956	0.9985	0.9990	0.9956	0.9995	0.9986	0.9973
24	0.9987	0.9994	0.9980	0.9992	0.9985	0.9998	0.9998	0.9996	0.9998	0.9990	0.9987	0.9984	0.9989	0.9982	0.9945	0.9966	0.9992	0.9995	0.9967	0.9974	0.9988	0.9994	0.9994	0.9991	0.9992	0.0000	0.9992	0.9994	0.9990	0.9988	0.9986	0.9986	0.9954	0.9992	0.9984	0.9969	0.9979	0.9987
25	0.9978	0.9997	0.9981	0.9987	0.9980	0.9987	0.9995	0.9985	0.9994	0.9995	0.9969	0.9989	0.9983	0.9983	0.9983	0.9971	0.9986	0.9991	0.9989	0.9998	0.9986	0.9993	0.9973	0.9991	0.9991	0.9992	0.0000	0.9991	0.9969	0.9981	0.9973	0.9987	0.9973	0.9979	0.9984	0.9998	0.9966
26	0.9990	0.9997	0.9990	0.9967	0.9989	0.9991	0.9992	0.9989	0.9991	0.9996	0.9983	0.9997	0.9993	0.9983	0.9983	0.9971	0.9986	0.9991	0.9989	0.9998	0.9986	0.9993	0.9973	0.9991	0.9991	0.0000	0.9972	0.9984	0.9984	0.9978	0.9993	0.9992	0.9974	0.9988	0.9988	0.9995	0.9984	0.9959
27	0.9999	0.9978	0.9993	0.9973	0.9984	0.9992	0.9996	0.9985	0.9960	0.9979	0.9987	0.9981	0.9991	0.9978	0.9988	0.9985	0.9995	0.9977	0.9991	0.9990	0.9994	0.9965	0.9997	0.9991	0.9972	0.0000	0.9989	0.9984	0.9993	0.9947	0.9994	0.9991	0.9942	0.9980	0.9983	0.9984	0.9982
28	0.9991	0.9997	0.9985	0.9957	0.9974	0.9990	0.9981	0.9983	0.9984	0.9989	0.9994	0.9994	0.9994	0.9989	0.9990	0.9992	0.9964	0.9991	0.9996	0.9982	0.9996	0.9991	0.9996	0.9994	0.9938	0.9989	0.0000	0.9997	0.9992	0.9973	0.9986	0.9984	0.9991	0.9989	0.9998	0.9990	0.9985
29	0.9987	0.9990	0.9976	0.9995	0.9994	0.9992	0.9980	0.9972	0.9994	0.9996	0.9995	0.9995	0.9955	0.9999	0.9983	0.9985	0.9999	0.9986	0.9988	0.9960	0.9996	0.9994	0.9990	0.9984	0.9969	0.9984	0.9984	0.9984	0.9997	0.0000	0.9981	0.9984	0.9991	0.9993	0.9985	0.9993	0.9993
30	0.9993	0.9986	0.9991	0.9980	0.9995	0.9980	0.9998	0.9995	0.9991	0.9989	0.9975	0.9959	0.9999	0.9999	0.9988	0.9986	0.9971	0.9990	0.9992	0.9998	0.9963	0.9991	0.9988	0.9981	0.9978	0.9993	0.9992	0.9978	0.9993	0.0000	0.9973	0.9959	0.9992	0.9988	0.9964	0.9977	0.9994	0.9978
31	0.9970	0.9996	0.9993	0.9980	0.9984	0.9996	0.9987	0.9988	0.9995	0.9992	0.9994	0.9993	0.9994	0.9984	0.9986	0.9999	0.9995	0.9997	0.9949	0.9964	0.9990	0.9994	0.9996	0.9986	0.9973	0.9993	0.9947	0.9973	0.9994	0.9973	0.0000	0.9991	0.9994	0.9985	0.9995	0.9992	0.9993	0.9984
32	0.9991	0.9996	0.9994	0.9972	0.9996	0.9987	0.9977	0.9988	0.9995	0.9992	0.9994	0.9997	0.9979	0.9985	0.9982	0.9991	0.9994	0.9992	0.9988	0.9994	0.9990	0.9978	0.9986	0.9977	0.9992	0.9994	0.9986	0.9975	0.9959	0.9991	0.0000	0.9987	0.9994	0.9985	0.9965	0.9979	0.9994	0.9998
33	0.9965	0.9996	0.9970	0.9993	0.9990	0.9989	0.9996	0.9991	0.9960	0.9979	0.9990	0.9964	0.9984	0.9959	0.9994	0.9980	0.9969	0.9994	0.9946	0.9992	0.9975	0.9994	0.9986	0.9977	0.9974	0.9991	0.9984	0.9974	0.9992	0.9994	0.9959	0.0000	0.9987	0.9981	0.9991	0.9998	0.9993	0.9995
34	0.9985	0.9999	0.9993	0.9997	0.9984	0.9994	0.9943	0.9983	0.9984	0.9989	0.9970	0.9979	0.9995	0.9996	0.9979	0.9988	0.9992	0.9989	0.9977	0.9990	0.9992	0.9973	0.9988	0.9981	0.9993	0.9947	0.9973	0.9984	0.9992	0.9984	0.9991	0.9994	0.9987	0.0000	0.9983	0.9993	0.9990	0.9977
35	0.9995	0.9993	0.9983	0.9998	0.9975	0.9987	0.9996	0.9987	0.9996	1.0000	0.9986	0.9984	0.9996	0.9995	0.9977	0.9978	0.9985	0.9991	0.9984	0.9982	0.9995	0.9961	0.9956	0.9984	0.9979	0.9988	0.9980	0.9989	0.9993	0.9964	0.9995	0.9965	0.9981	0.9983	0.0000	0.9998	0.9983	0.9975
36	0.9996	0.9969	0.9996	0.9959	0.9984	0.9997	0.9975	0.9986	0.9974	0.9982	0.9996	0.9980	0.9994	0.9996	0.9998	0.9995	0.9992	0.9974	0.9983	0.9988	0.9989	0.9995	0.9995	0.9989	0.9984	0.9995	0.9983	0.9968	0.9985	0.9977	0.9992	0.9979	0.9991	0.9993	0.9998	0.0000	0.9990	0.9979
37	0.9994	0.9940	0.9995	0.9994	0.9977	0.9984	0.9993	0.9999	0.9989	0.9991	0.9977	0.9991	0.9967	0.9958	0.9978	0.9980	0.9969	0.9994	0.9946	0.9992	0.9975	0.9994	0.9986	0.9979	0.9998	0.9984	0.9990	0.9993	0.9994	0.9993	0.9994	0.9998	0.9990	0.9993	0.9990	0.0000	0.9990	0.9995
38	0.9998	0.9942	0.9990	0.9995	0.9988	0.9996	0.9990	0.9987	0.9983	0.9990	0.9982	0.9996	0.9994	0.9982	0.9992	0.9992	0.9978	0.9977	0.9985	0.9976	0.9993	0.9941	0.9973	0.9987	0.9996	0.9959	0.9982	0.9985	0.9993	0.9978	0.9984	0.9998	0.9989	0.9977	0.9975	0.9979	0.9995	0.0000

Appendix C: Changes Between Performance Measures

Table C.1: Changes in BCT accuracy between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0060	0.0040	0.0060	0.0090	0.0030	0.0020	0.0180	-0.0030	-0.0030	0.0180	0.0180	0.0120	0.0160	0.0060	0.0120	0.0100	0.0130	0.0170	0.0060	0.0140
3	0.0280	0.0340	0.0350	0.0330	0.0330	0.0320	0.0440	0.0320	0.0240	0.0370	0.0430	0.0400	0.0410	0.0300	0.0430	0.0380	0.0340	0.0460	0.0370	0.0390
4	0.0300	0.0340	0.0350	0.0340	0.0320	0.0310	0.0420	0.0320	0.0210	0.0360	0.0440	0.0400	0.0400	0.0290	0.0400	0.0410	0.0420	0.0450	0.0360	0.0360
5	0.0300	0.0310	0.0350	0.0320	0.0320	0.0280	0.0420	0.0320	0.0220	0.0350	0.0430	0.0410	0.0370	0.0310	0.0390	0.0390	0.0410	0.0450	0.0360	0.0360
6	0.0280	0.0300	0.0340	0.0320	0.0230	0.0270	0.0380	0.0300	0.0170	0.0340	0.0400	0.0400	0.0380	0.0240	0.0390	0.0350	0.0380	0.0390	0.0310	0.0360
7	0.0310	0.0310	0.0320	0.0370	0.0300	0.0330	0.0440	0.0300	0.0210	0.0360	0.0400	0.0390	0.0390	0.0300	0.0390	0.0350	0.0390	0.0440	0.0350	0.0370
8	0.0290	0.0300	0.0330	0.0330	0.0280	0.0270	0.0390	0.0300	0.0190	0.0350	0.0430	0.0390	0.0370	0.0280	0.0390	0.0370	0.0390	0.0440	0.0360	0.0370
9	0.0040	0.0050	0.0130	0.0090	0.0100	0.0050	0.0190	0.0060	-0.0090	0.0120	0.0190	0.0150	0.0130	0.0120	0.0140	0.0140	0.0120	0.0210	0.0050	0.0070
10	-0.0050	-0.0030	-0.0020	0.0020	-0.0040	-0.0100	0.0090	-0.0060	-0.0160	0.0030	0.0010	0.0110	0.0090	-0.0140	0.0030	0.0040	0.0050	0.0120	0.0010	0.0060
11	0.0040	0.0120	0.0070	0.0080	0.0050	0.0020	0.0230	0.0080	-0.0030	-0.0010	0.0180	0.0130	0.0170	0.0070	0.0190	0.0100	0.0140	0.0170	0.0150	0.0170
12	0.0100	0.0100	0.0150	0.0160	0.0130	0.0110	0.0210	0.0140	0.0060	0.0130	0.0180	0.0210	0.0180	0.0110	0.0270	0.0230	0.0210	0.0240	0.0180	0.0230
13	-0.0100	-0.0120	-0.0010	-0.0050	-0.0160	-0.0040	0.0010	-0.0090	-0.0170	-0.0040	0.0030	0.0030	0.0000	-0.0100	0.0000	0.0010	-0.0040	0.0100	-0.0100	-0.0070
14	-0.0130	-0.0060	-0.0010	-0.0070	-0.0070	-0.0040	-0.0010	-0.0090	-0.0180	-0.0050	-0.0010	-0.0010	-0.0090	-0.0100	0.0020	-0.0090	-0.0020	0.0080	-0.0090	-0.0080
15	-0.0040	0.0000	-0.0050	-0.0110	-0.0030	-0.0070	0.0120	0.0010	-0.0110	0.0010	-0.0010	0.0000	-0.0010	-0.0090	0.0110	0.0040	0.0080	0.0060	0.0070	0.0030
16	-0.0200	-0.0260	-0.0260	-0.0150	-0.0240	-0.0240	-0.0180	-0.0230	-0.0360	-0.0250	-0.0230	-0.0260	-0.0210	-0.0260	-0.0210	-0.0220	-0.0220	-0.0110	-0.0270	-0.0260
17	-0.0040	0.0020	0.0070	0.0070	0.0050	0.0030	0.0130	0.0010	-0.0090	0.0060	0.0160	0.0170	0.0120	-0.0030	0.0110	0.0090	0.0130	0.0180	0.0090	0.0050
18	-0.0160	-0.0200	-0.0110	-0.0150	-0.0160	-0.0120	0.0000	-0.0130	-0.0210	-0.0010	0.0060	-0.0020	-0.0150	-0.0180	0.0030	0.0010	-0.0040	0.0000	-0.0030	-0.0050
19	0.0150	0.0210	0.0220	0.0200	0.0210	0.0120	0.0290	0.0190	0.0080	0.0240	0.0300	0.0260	0.0290	0.0170	0.0280	0.0200	0.0270	0.0320	0.0210	0.0260
20	-0.0020	-0.0030	0.0070	0.0070	0.0000	0.0030	0.0060	0.0040	-0.0060	0.0090	0.0120	0.0100	0.0070	0.0050	0.0140	0.0040	0.0090	0.0140	0.0060	0.0030
21	-0.0180	-0.0200	-0.0050	-0.0140	-0.0130	-0.0200	0.0020	-0.0130	-0.0270	-0.0090	-0.0100	-0.0040	-0.0080	-0.0200	-0.0070	-0.0030	-0.0090	-0.0030	-0.0130	-0.0150
22	-0.0250	-0.0270	-0.0210	-0.0170	-0.0240	-0.0220	-0.0080	-0.0330	-0.0320	-0.0230	-0.0140	-0.0080	-0.0190	-0.0290	-0.0080	-0.0140	-0.0130	-0.0090	-0.0180	-0.0170
23	-0.0280	-0.0330	-0.0320	-0.0390	-0.0280	-0.0370	-0.0310	-0.0370	-0.0460	-0.0230	-0.0250	-0.0340	-0.0190	-0.0350	-0.0350	-0.0270	-0.0360	-0.0200	-0.0310	-0.0290
24	-0.0260	-0.0210	-0.0280	-0.0190	-0.0250	-0.0260	-0.0210	-0.0280	-0.0390	-0.0190	-0.0130	-0.0180	-0.0180	-0.0290	-0.0140	-0.0270	-0.0050	-0.0150	-0.0220	-0.0160
25	0.0050	0.0040	-0.0010	0.0030	-0.0020	0.0010	0.0020	0.0010	-0.0120	0.0060	0.0160	0.0070	0.0100	-0.0020	0.0080	0.0050	0.0060	0.0130	0.0010	0.0030
26	-0.0380	-0.0380	-0.0480	-0.0320	-0.0530	-0.0440	-0.0330	-0.0460	-0.0500	-0.0420	-0.0240	-0.0430	-0.0320	-0.0500	-0.0420	-0.0380	-0.0390	-0.0340	-0.0390	-0.0320
27	0.0310	0.0340	0.0360	0.0350	0.0330	0.0320	0.0450	0.0330	0.0230	0.0380	0.0460	0.0420	0.0420	0.0320	0.0420	0.0400	0.0420	0.0460	0.0370	0.0390
28	0.0320	0.0350	0.0350	0.0360	0.0340	0.0330	0.0440	0.0340	0.0230	0.0390	0.0430	0.0430	0.0410	0.0330	0.0430	0.0410	0.0430	0.0460	0.0380	0.0400
29	0.0330	0.0350	0.0380	0.0370	0.0350	0.0340	0.0450	0.0350	0.0250	0.0400	0.0460	0.0440	0.0420	0.0340	0.0440	0.0420	0.0440	0.0480	0.0390	0.0410
30	0.0320	0.0360	0.0370	0.0360	0.0340	0.0330	0.0440	0.0340	0.0240	0.0390	0.0450	0.0430	0.0410	0.0330	0.0430	0.0410	0.0430	0.0470	0.0390	0.0400
31	0.0330	0.0350	0.0380	0.0370	0.0350	0.0340	0.0450	0.0340	0.0250	0.0400	0.0460	0.0430	0.0420	0.0340	0.0440	0.0420	0.0440	0.0470	0.0390	0.0410
32	0.0330	0.0360	0.0380	0.0370	0.0350	0.0340	0.0450	0.0350	0.0250	0.0390	0.0460	0.0440	0.0420	0.0340	0.0440	0.0420	0.0430	0.0480	0.0390	0.0410
33	-0.0030	0.0020	0.0090	0.0070	0.0020	-0.0010	0.0060	0.0040	-0.0060	0.0090	0.0120	0.0110	0.0190	0.0010	0.0130	0.0060	0.0090	0.0120	0.0050	0.0100
34	-0.0090	-0.0100	-0.0080	-0.0080	-0.0020	-0.0120	-0.0020	-0.0170	-0.0150	-0.0020	-0.0110	-0.0010	-0.0080	-0.0080	-0.0120	0.0010	-0.0070	0.0020	-0.0070	0.0000
35	0.0120	0.0110	0.0130	0.0160	0.0100	0.0140	0.0180	0.0120	0.0000	0.0200	0.0200	0.0200	0.0180	0.0150	0.0230	0.0210	0.0170	0.0240	0.0190	0.0180
36	-0.0010	0.0020	0.0070	0.0040	-0.0010	0.0050	0.0150	0.0050	-0.0060	0.0140	0.0120	0.0120	0.0070	0.0040	0.0160	0.0080	0.0120	0.0130	0.0050	0.0060
37	0.0170	0.0220	0.0210	0.0220	0.0210	0.0190	0.0300	0.0190	0.0090	0.0200	0.0310	0.0240	0.0260	0.0160	0.0260	0.0250	0.0300	0.0290	0.0240	0.0260
38	0.0130	0.0040	-0.0000	0.0040	0.0070	0.0030	0.0180	0.0090	-0.0080	0.0050	0.0080	0.0080	0.0080	-0.0010	0.0120	0.0110	0.0100	0.0120	0.0090	0.0110

Table C.2: Changes in BCT TPR between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-0.0006	-0.0078	-0.0075	0.0072	-0.0073	-0.0073	0.0038	-0.0136	-0.0001	0.0025	-0.0047	-0.0018	-0.0021	0.0051	0.0018	0.0031	-0.0034	0.0013	-0.0078	0.0007
3	0.0174	0.0191	0.0214	0.0253	0.0195	0.0161	0.0234	0.0204	0.0195	0.0188	0.0188	0.0253	0.0247	0.0230	0.0286	0.0249	0.0133	0.0299	0.0191	0.0204
4	0.0208	0.0208	0.0247	0.0286	0.0195	0.0181	0.0234	0.0221	0.0181	0.0207	0.0221	0.0273	0.0247	0.0233	0.0273	0.0299	0.0234	0.0299	0.0208	0.0221
5	0.0182	0.0129	0.0221	0.0207	0.0169	0.0116	0.0208	0.0195	0.0169	0.0142	0.0195	0.0260	0.0168	0.0221	0.0207	0.0273	0.0208	0.0273	0.0182	0.0142
6	0.0197	0.0197	0.0236	0.0275	0.0172	0.0160	0.0223	0.0210	0.0172	0.0210	0.0198	0.0275	0.0236	0.0236	0.0275	0.0288	0.0223	0.0264	0.0197	0.0210
7	0.0185	0.0174	0.0202	0.0275	0.0149	0.0172	0.0211	0.0164	0.0161	0.0187	0.0187	0.0241	0.0213	0.0213	0.0252	0.0242	0.0200	0.0277	0.0163	0.0198
8	0.0142	0.0142	0.0181	0.0253	0.0096	0.0096	0.0168	0.0155	0.0096	0.0155	0.0188	0.0220	0.0181	0.0181	0.0220	0.0266	0.0168	0.0233	0.0175	0.0188
9	0.0080	0.0080	0.0151	0.0180	0.0099	0.0121	0.0160	0.0104	0.0056	0.0125	0.0093	0.0158	0.0141	0.0173	0.0169	0.0203	0.0085	0.0203	0.0069	0.0082
10	-0.0759	-0.0529	-0.0766	-0.0589	-0.0726	-0.0957	-0.0411	-0.0608	-0.0865	-0.0562	-0.0977	-0.0267	-0.0536	-0.0951	-0.0681	-0.0807	-0.0733	-0.0714	-0.0621	-0.0654
11	0.0065	0.0116	0.0094	0.0144	0.0022	0.0022	0.0153	0.0078	0.0093	-0.0013	0.0078	0.0144	0.0166	0.0125	0.0164	0.0126	0.0092	0.0126	0.0127	0.0119
12	-0.2341	-0.2733	-0.2890	-0.2459	-0.2158	-0.1962	-0.2119	-0.1739	-0.1177	-0.2524	-0.2328	-0.2066	-0.2302	-0.1713	-0.0890	-0.1661	-0.1922	-0.2053	-0.1556	-0.1347
13	-0.0036	-0.0083	0.0026	0.0030	-0.0096	-0.0026	0.0001	0.0012	0.0020	-0.0035	0.0023	0.0077	0.0026	0.0038	0.0018	0.0055	0.0001	0.0113	-0.0048	-0.0012
14	-0.0273	-0.0353	-0.0234	-0.0248	-0.0393	-0.0366	-0.0274	-0.0287	-0.0286	-0.0287	-0.0394	-0.0195	-0.0314	-0.0314	-0.0248	-0.0369	-0.0327	-0.0155	-0.0460	-0.0367
15	0.0009	0.0020	0.0024	0.0051	0.0031	-0.0063	0.0070	0.0057	-0.0028	0.0022	-0.0037	0.0122	0.0024	0.0001	0.0134	0.0100	0.0093	0.0076	0.0056	0.0010
16	-0.0349	-0.0465	-0.0541	-0.0317	-0.0431	-0.0338	-0.0485	-0.0475	-0.0547	-0.0475	-0.0637	-0.0549	-0.0402	-0.0402	-0.0433	-0.0443	-0.0601	-0.0443	-0.0604	-0.0637
17	0.0046	0.0025	0.0118	0.0135	0.0055	0.0033	0.0062	0.0049	0.0044	0.0027	0.0049	0.0178	0.0129	0.0021	0.0103	0.0137	0.0072	0.0159	0.0068	0.0016
18	-0.0903	-0.0944	-0.0699	-0.0742	-0.0875	-0.0628	-0.0589	-0.0807	-0.0710	-0.0602	-0.0519	-0.0495	-0.0946	-0.0946	-0.0537	-0.0565	-0.0548	-0.0647	-0.0779	-0.0643
19	0.0115	0.0146	0.0175	0.0193	0.0143	0.0060	0.0162	0.0149	0.0112	0.0149	0.0149	0.0193	0.0206	0.0164	0.0224	0.0175	0.0151	0.0237	0.0125	0.0169
20	-0.1524	-0.1366	-0.1249	-0.1210	-0.1143	-0.1301	-0.1498	-0.1038	-0.1143	-0.1038	-0.1353	-0.1210	-0.1170	-0.1249	-0.0737	-0.1275	-0.1340	-0.1275	-0.1052	-0.1353
21	-0.0094	-0.0071	0.0015	0.0007	-0.0154	-0.0131	0.0025	-0.0058	-0.0084	-0.0046	-0.0105	0.0007	-0.0067	-0.0090	-0.0005	0.0078	-0.0045	-0.0015	-0.0106	-0.0116
22	-0.0513	-0.0538	-0.0499	-0.0261	-0.0601	-0.0501	-0.0437	-0.0600	-0.0551	-0.0550	-0.0525	-0.0261	-0.0623	-0.0524	-0.0410	-0.0397	-0.0363	-0.0347	-0.0463	-0.0600
23	-0.0500	-0.0543	-0.0568	-0.0593	-0.0448	-0.0642	-0.0581	-0.0573	-0.0620	-0.0487	-0.0530	-0.0615	-0.0418	-0.0439	-0.0615	-0.0387	-0.0753	-0.0452	-0.0543	-0.0551
24	-0.0094	-0.0058	-0.0019	-0.0040	-0.0107	-0.0155	-0.0128	-0.0105	-0.0107	-0.0093	-0.0105	-0.0040	-0.0043	-0.0091	-0.0028	-0.0051	-0.0020	-0.0039	-0.0142	-0.0045
25	0.0058	0.0005	0.0012	0.0094	0.0013	0.0035	0.0041	0.0071	-0.0019	0.0028	0.0093	0.0104	0.0108	0.0044	0.0104	0.0117	0.0031	0.0128	0.0005	0.0007
26	-0.1145	-0.1333	-0.1332	-0.1029	-0.1271	-0.1421	-0.1232	-0.1282	-0.1196	-0.1395	-0.1132	-0.1330	-0.1181	-0.1407	-0.1292	-0.1167	-0.1269	-0.1279	-0.1371	-0.1207
27	0.0196	0.0196	0.0235	0.0274	0.0183	0.0183	0.0234	0.0209	0.0183	0.0209	0.0221	0.0274	0.0247	0.0235	0.0274	0.0287	0.0222	0.0287	0.0196	0.0209
28	0.0189	0.0189	0.0190	0.0267	0.0176	0.0176	0.0215	0.0202	0.0138	0.0202	0.0164	0.0267	0.0228	0.0228	0.0267	0.0280	0.0215	0.0280	0.0189	0.0202
29	0.0208	0.0194	0.0247	0.0286	0.0195	0.0195	0.0234	0.0221	0.0195	0.0221	0.0221	0.0286	0.0247	0.0247	0.0286	0.0299	0.0234	0.0299	0.0208	0.0221
30	0.0194	0.0208	0.0233	0.0272	0.0181	0.0181	0.0220	0.0207	0.0181	0.0207	0.0207	0.0272	0.0233	0.0233	0.0272	0.0285	0.0220	0.0285	0.0208	0.0207
31	0.0208	0.0198	0.0247	0.0286	0.0195	0.0195	0.0234	0.0221	0.0195	0.0221	0.0221	0.0286	0.0247	0.0247	0.0286	0.0299	0.0234	0.0299	0.0208	0.0221
32	0.0208	0.0208	0.0247	0.0286	0.0195	0.0195	0.0234	0.0221	0.0195	0.0221	0.0221	0.0286	0.0247	0.0247	0.0286	0.0299	0.0234	0.0299	0.0208	0.0221
33	-0.0019	-0.0019	0.0109	0.0097	0.0006	-0.0045	0.0007	0.0045	0.0006	0.0057	0.0007	0.0072	0.0109	0.0045	0.0097	0.0072	0.0007	0.0060	0.0019	0.0019
34	-0.0129	-0.0235	-0.0280	-0.0199	-0.0142	-0.0311	-0.0251	-0.0285	-0.0121	-0.0222	-0.0412	-0.0199	-0.0217	-0.0090	-0.0346	-0.0122	-0.0272	-0.0080	-0.0277	-0.0095
35	0.0090	0.0068	0.0118	0.0157	0.0077	0.0066	0.0083	0.0113	0.0034	0.0103	0.0049	0.0103	0.0107	0.0161	0.0168	0.0181	0.0073	0.0181	0.0100	0.0113
36	-0.0583	-0.0583	-0.0419	-0.0339	-0.0721	-0.0555	-0.0391	-0.0445	-0.0388	-0.0487	-0.0487	-0.0422	-0.0544	-0.0378	-0.0422	-0.0534	-0.0349	-0.0367	-0.0542	-0.0570
37	0.0156	0.0156	0.0164	0.0214	0.0123	0.0123	0.0162	0.0138	0.0133	0.0138	0.0149	0.0203	0.0185	0.0154	0.0193	0.0227	0.0172	0.0206	0.0146	0.0159
38	-0.0713	-0.1173	-0.1003	-0.0964	-0.0792	-0.0923	-0.0818	-0.0765	-0.1120	-0.0963	-0.1160	-0.1161	-0.1003	-0.1200	-0.0898	-0.0687	-0.1213	-0.0885	-0.1042	-0.0765

Table C.3: Changes in BCT FPR between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-0.0413	-0.0520	-0.0522	-0.0304	-0.0497	-0.0454	-0.0821	-0.0388	-0.0066	-0.0762	-0.0976	-0.0629	-0.0762	-0.0281	-0.0583	-0.0454	-0.0778	-0.0758	-0.0626	-0.0715
3	-0.0678	-0.0856	-0.0813	-0.0616	-0.0832	-0.0838	-0.1158	-0.0746	-0.0425	-0.0986	-0.1244	-0.0918	-0.0986	-0.0592	-0.0942	-0.0813	-0.1042	-0.1047	-0.0986	-0.1029
4	-0.0634	-0.0798	-0.0720	-0.0547	-0.0763	-0.0755	-0.1065	-0.0677	-0.0332	-0.0892	-0.1186	-0.0849	-0.0928	-0.0512	-0.0849	-0.0791	-0.1057	-0.0978	-0.0892	-0.0864
5	-0.0728	-0.0857	-0.0814	-0.0641	-0.0857	-0.0798	-0.1159	-0.0771	-0.0426	-0.0986	-0.1245	-0.0943	-0.0986	-0.0641	-0.0943	-0.0814	-0.1116	-0.1072	-0.0986	-0.1029
6	-0.0447	-0.0510	-0.0599	-0.0361	-0.0181	-0.0533	-0.0746	-0.0490	-0.0014	-0.0640	-0.0964	-0.0728	-0.0771	-0.0032	-0.0662	-0.0402	-0.0769	-0.0660	-0.0508	-0.0749
7	-0.0697	-0.0669	-0.0626	-0.0690	-0.0748	-0.0862	-0.1207	-0.0740	-0.0317	-0.0877	-0.0978	-0.0834	-0.0956	-0.0532	-0.0755	-0.0626	-0.0928	-0.0884	-0.0956	-0.0841
8	-0.0733	-0.0834	-0.0805	-0.0632	-0.0834	-0.0791	-0.1135	-0.0762	-0.0417	-0.0977	-0.1250	-0.0934	-0.0977	-0.0618	-0.0934	-0.0791	-0.1107	-0.1078	-0.0992	-0.1020
9	0.2037	0.2220	0.1794	0.2279	0.1751	0.2732	0.1918	0.2150	0.2963	0.2091	0.1207	0.1821	0.2091	0.1810	0.2134	0.2263	0.1805	0.1848	0.2403	0.2360
10	-0.0546	-0.0599	-0.0619	-0.0473	-0.0650	-0.0607	-0.0913	-0.0512	-0.0231	-0.0766	-0.1038	-0.0710	-0.0817	-0.0396	-0.0723	-0.0670	-0.0921	-0.0929	-0.0766	-0.0860
11	0.7224	0.7095	0.7638	0.7310	0.6095	0.7138	0.6293	0.6181	0.9026	0.8466	0.6207	0.8009	0.7966	0.7310	0.6009	0.7138	0.7336	0.6379	0.7466	0.6422
12	-0.0660	-0.0779	-0.0778	-0.0605	-0.0789	-0.0725	-0.1070	-0.0693	-0.0337	-0.0887	-0.1125	-0.0865	-0.0908	-0.0542	-0.0865	-0.0757	-0.1027	-0.0984	-0.0897	-0.0962
13	0.0844	0.0785	0.0617	0.0789	0.0996	0.0546	0.0554	0.1082	0.1498	0.0585	0.0608	0.0699	0.0656	0.1212	0.0558	0.0617	0.0878	0.0499	0.0937	0.0965
14	-0.0313	-0.0554	-0.0511	-0.0290	-0.0570	-0.0575	-0.0760	-0.0404	-0.0059	-0.0603	-0.0894	-0.0544	-0.0539	-0.0306	-0.0624	-0.0431	-0.0748	-0.0737	-0.0651	-0.0630
15	0.0634	0.0504	0.0816	0.1257	0.0773	0.0480	0.0135	0.0590	0.0734	0.0509	0.0452	0.1089	0.0643	0.0854	0.0418	0.0614	0.0514	0.0490	0.0308	0.0332
16	-0.0249	-0.0308	-0.0317	-0.0215	-0.0325	-0.0229	-0.0627	-0.0309	0.0053	-0.0402	-0.0713	-0.0376	-0.0402	-0.0110	-0.0376	-0.0282	-0.0619	-0.0629	-0.0472	-0.0533
17	0.2332	0.1527	0.1841	0.1608	0.1527	0.1435	0.1090	0.1749	0.2364	0.1263	0.0734	0.1441	0.1668	0.1608	0.1306	0.1705	0.1134	0.1312	0.1398	0.1355
18	-0.0472	-0.0522	-0.0505	-0.0320	-0.0562	-0.0505	-0.0863	-0.0502	-0.0144	-0.0744	-0.0989	-0.0621	-0.0651	-0.0373	-0.0701	-0.0585	-0.0767	-0.0777	-0.0783	-0.0734
19	0.1788	0.1659	0.1702	0.1618	0.1659	0.1702	0.1357	0.1745	0.2090	0.1530	0.1271	0.1573	0.1530	0.1874	0.1829	0.1958	0.1400	0.1700	0.1786	0.1743
20	-0.0615	-0.0676	-0.0713	-0.0552	-0.0688	-0.0713	-0.1001	-0.0636	-0.0302	-0.0851	-0.1121	-0.0808	-0.0828	-0.0564	-0.0785	-0.0644	-0.0981	-0.0949	-0.0828	-0.0860
21	0.1068	0.1435	0.0840	0.1296	0.0442	0.1053	0.0637	0.0954	0.1583	0.0880	0.0764	0.0782	0.0668	0.1154	0.0924	0.1053	0.0964	0.0652	0.0809	0.0908
22	-0.0274	-0.0337	-0.0360	-0.0138	-0.0437	-0.0377	-0.0755	-0.0217	-0.0006	-0.0483	-0.0775	-0.0473	-0.0583	-0.0138	-0.0573	-0.0377	-0.0595	-0.0586	-0.0516	-0.0643
23	-0.0233	-0.0250	-0.0244	-0.0015	-0.0268	-0.0244	-0.0477	-0.0145	0.0163	-0.0454	-0.0600	-0.0298	-0.0454	0.0022	-0.0280	-0.0150	-0.0508	-0.0484	-0.0360	-0.0422
24	0.1247	0.1176	0.1739	0.1044	0.1176	0.0988	0.0932	0.1320	0.1838	0.0931	0.0614	0.1090	0.1104	0.1391	0.0916	0.1508	0.0513	0.0960	0.0873	0.1003
25	0.1463	0.1184	0.1824	0.1848	0.2229	0.1974	0.2674	0.2315	0.2212	0.1503	0.1543	0.2143	0.1951	0.1997	0.1994	0.2272	0.1821	0.1864	0.1951	0.1758
26	-0.0285	-0.0442	-0.0249	-0.0213	-0.0224	-0.0372	-0.0662	-0.0247	0.0057	-0.0490	-0.0816	-0.0378	-0.0530	-0.0131	-0.0378	-0.0290	-0.0564	-0.0562	-0.0530	-0.0587
27	-0.0673	-0.0803	-0.0760	-0.0587	-0.0803	-0.0760	-0.1156	-0.0716	-0.0372	-0.0932	-0.1242	-0.0889	-0.0983	-0.0587	-0.0889	-0.0760	-0.1061	-0.1018	-0.0932	-0.0975
28	-0.0755	-0.0884	-0.0841	-0.0669	-0.0884	-0.0841	-0.1186	-0.0798	-0.0474	-0.1014	-0.1272	-0.0970	-0.1014	-0.0669	-0.0970	-0.0841	-0.1143	-0.1079	-0.1014	-0.1057
29	-0.0745	-0.0874	-0.0831	-0.0659	-0.0874	-0.0831	-0.1176	-0.0788	-0.0443	-0.1004	-0.1262	-0.0960	-0.1004	-0.0659	-0.0960	-0.0831	-0.1133	-0.1090	-0.1004	-0.1047
30	-0.0745	-0.0874	-0.0831	-0.0659	-0.0874	-0.0831	-0.1176	-0.0788	-0.0443	-0.1004	-0.1262	-0.0961	-0.1004	-0.0659	-0.0961	-0.0831	-0.1133	-0.1090	-0.1004	-0.1047
31	-0.0612	-0.0741	-0.0698	-0.0526	-0.0741	-0.0698	-0.1043	-0.0491	-0.0310	-0.0871	-0.1129	-0.0664	-0.0871	-0.0526	-0.0827	-0.0698	-0.1000	-0.0793	-0.0871	-0.0914
32	-0.0764	-0.0893	-0.0850	-0.0677	-0.0893	-0.0850	-0.1195	-0.0807	-0.0462	-0.1010	-0.1281	-0.0979	-0.1022	-0.0677	-0.0979	-0.0850	-0.1139	-0.1108	-0.1022	-0.1065
33	0.0138	-0.0088	0.0051	0.0080	0.0008	-0.0045	-0.0149	0.0046	0.0343	-0.0121	-0.0428	-0.0174	-0.0409	0.0176	-0.0174	0.0051	-0.0298	-0.0255	-0.0073	-0.0308
34	-0.0263	-0.0411	-0.0444	-0.0252	-0.0487	-0.0425	-0.0732	-0.0268	0.0020	-0.0616	-0.0761	-0.0554	-0.0483	-0.0176	-0.0478	-0.0444	-0.0631	-0.0569	-0.0578	-0.0564
35	0.0731	0.0876	0.1056	0.0680	0.1150	0.0371	0.0711	0.1099	0.1033	0.0335	0.0214	0.0105	0.0609	0.0954	0.0515	0.0645	0.0617	0.0797	0.0472	0.0840
36	-0.0565	-0.0695	-0.0652	-0.0440	-0.0708	-0.0704	-0.0996	-0.0622	-0.0237	-0.0903	-0.1056	-0.0781	-0.0811	-0.0479	-0.0833	-0.0665	-0.0914	-0.0858	-0.0811	-0.0854
37	0.2557	0.1873	0.1916	0.1810	0.1317	0.1638	0.1293	0.1681	0.2581	0.2577	0.1207	0.2620	0.2021	0.2088	0.1786	0.2193	0.1336	0.1935	0.1743	0.1700
38	-0.0693	-0.0764	-0.0626	-0.0513	-0.0740	-0.0685	-0.1065	-0.0677	-0.0309	-0.0822	-0.1081	-0.0814	-0.0846	-0.0525	-0.0814	-0.0662	-0.1010	-0.0897	-0.0893	-0.0889

Table C.4: Changes in NB accuracy between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-0.0070	-0.0110	-0.0110	-0.0110	-0.0110	-0.0120	-0.0100	-0.0090	-0.0080	-0.0090	-0.0060	-0.0100	-0.0050	-0.0110	-0.0030	-0.0090	-0.0040	-0.0160	-0.0120	-0.0090
3	0.1110	0.1170	0.1120	0.1130	0.1120	0.1100	0.1140	0.1130	0.1110	0.1150	0.1120	0.1140	0.1120	0.1130	0.1170	0.1070	0.1180	0.1030	0.1110	0.1100
4	0.0360	0.0370	0.0440	0.0310	0.0370	0.0370	0.0350	0.0400	0.0360	0.0400	0.0390	0.0410	0.0390	0.0380	0.0440	0.0370	0.0450	0.0370	0.0350	0.0410
5	0.1330	0.1350	0.1360	0.1340	0.1290	0.1330	0.1340	0.1380	0.1330	0.1340	0.1400	0.1330	0.1370	0.1360	0.1380	0.1320	0.1420	0.1360	0.1380	0.1330
6	0.0740	0.0750	0.0740	0.0760	0.0730	0.0770	0.0750	0.0810	0.0760	0.0770	0.0730	0.0720	0.0770	0.0760	0.0830	0.0680	0.0800	0.0720	0.0800	0.0770
7	0.0930	0.0940	0.0940	0.0930	0.0950	0.0920	0.0960	0.0970	0.0970	0.0960	0.0970	0.0970	0.0950	0.0980	0.1030	0.0930	0.0990	0.0920	0.0950	0.0960
8	0.1160	0.1120	0.1110	0.1130	0.1090	0.1130	0.1160	0.1220	0.1120	0.1210	0.1140	0.1170	0.1160	0.1130	0.1230	0.1160	0.1240	0.1130	0.1220	0.1160
9	0.1310	0.1320	0.1320	0.1300	0.1290	0.1290	0.1320	0.1330	0.1310	0.1320	0.1330	0.1310	0.1330	0.1330	0.1370	0.1310	0.1350	0.1280	0.1310	0.1300
10	0.0720	0.0730	0.0720	0.0710	0.0700	0.0710	0.0720	0.0790	0.0700	0.0770	0.0680	0.0750	0.0740	0.0720	0.0800	0.0680	0.0650	0.0670	0.0740	0.0760
11	0.1750	0.1760	0.1760	0.1740	0.1730	0.1730	0.1760	0.1770	0.1750	0.1760	0.1770	0.1750	0.1770	0.1770	0.1810	0.1750	0.1790	0.1720	0.1750	0.1740
12	0.1440	0.1450	0.1450	0.1430	0.1420	0.1420	0.1450	0.1460	0.1440	0.1450	0.1460	0.1440	0.1460	0.1460	0.1500	0.1440	0.1480	0.1410	0.1440	0.1430
13	0.0530	0.0540	0.0540	0.0520	0.0510	0.0510	0.0540	0.0550	0.0530	0.0540	0.0550	0.0530	0.0550	0.0550	0.0590	0.0530	0.0570	0.0500	0.0530	0.0520
14	0.0490	0.0550	0.0450	0.0520	0.0480	0.0470	0.0510	0.0520	0.0500	0.0520	0.0510	0.0540	0.0500	0.0570	0.0560	0.0500	0.0540	0.0480	0.0480	0.0530
15	0.0930	0.0950	0.0970	0.0970	0.0940	0.0930	0.0950	0.0960	0.0940	0.0940	0.0990	0.0960	0.0980	0.0970	0.0990	0.0960	0.0990	0.0910	0.0910	0.0960
16	0.0400	0.0420	0.0380	0.0380	0.0390	0.0390	0.0450	0.0410	0.0410	0.0440	0.0410	0.0420	0.0390	0.0460	0.0500	0.0420	0.0470	0.0370	0.0400	0.0440
17	0.1410	0.1410	0.1420	0.1350	0.1390	0.1360	0.1410	0.1390	0.1400	0.1370	0.1410	0.1410	0.1420	0.1420	0.1440	0.1400	0.1420	0.1370	0.1380	0.1370
18	0.1290	0.1300	0.1320	0.1290	0.1290	0.1250	0.1290	0.1310	0.1270	0.1290	0.1300	0.1320	0.1310	0.1280	0.1340	0.1310	0.1310	0.1240	0.1280	0.1270
19	0.1560	0.1580	0.1580	0.1560	0.1540	0.1540	0.1570	0.1580	0.1560	0.1570	0.1580	0.1560	0.1590	0.1580	0.1620	0.1580	0.1600	0.1530	0.1560	0.1550
20	0.1460	0.1460	0.1490	0.1460	0.1450	0.1450	0.1470	0.1500	0.1490	0.1480	0.1480	0.1460	0.1430	0.1490	0.1550	0.1470	0.1480	0.1440	0.1450	0.1440
21	0.0540	0.0550	0.0550	0.0530	0.0520	0.0520	0.0550	0.0560	0.0540	0.0550	0.0560	0.0540	0.0560	0.0560	0.0600	0.0540	0.0580	0.0510	0.0540	0.0530
22	0.0590	0.0610	0.0590	0.0610	0.0570	0.0600	0.0670	0.0590	0.0630	0.0650	0.0610	0.0590	0.0620	0.0620	0.0650	0.0580	0.0660	0.0570	0.0610	0.0550
23	-0.1100	-0.1050	-0.1120	-0.1080	-0.1080	-0.1080	-0.1060	-0.1020	-0.1100	-0.1020	-0.1030	-0.1140	-0.1090	-0.1110	-0.0990	-0.1110	-0.1100	-0.1130	-0.1100	-0.1120
24	0.0230	0.0240	0.0230	0.0230	0.0210	0.0210	0.0240	0.0250	0.0250	0.0250	0.0260	0.0230	0.0240	0.0250	0.0300	0.0240	0.0270	0.0210	0.0230	0.0220
25	0.1280	0.1290	0.1290	0.1270	0.1260	0.1260	0.1290	0.1300	0.1280	0.1290	0.1300	0.1280	0.1300	0.1300	0.1340	0.1280	0.1320	0.1250	0.1280	0.1270
26	-0.0710	-0.0700	-0.0700	-0.0720	-0.0730	-0.0730	-0.0700	-0.0690	-0.0710	-0.0700	-0.0690	-0.0710	-0.0690	-0.0690	-0.0650	-0.0710	-0.0670	-0.0740	-0.0710	-0.0720
27	0.1540	0.1520	0.1550	0.1500	0.1510	0.1490	0.1540	0.1530	0.1510	0.1540	0.1470	0.1530	0.1550	0.1500	0.1580	0.1490	0.1550	0.1510	0.1470	0.1500
28	0.1570	0.1580	0.1570	0.1550	0.1560	0.1520	0.1590	0.1580	0.1570	0.1570	0.1580	0.1570	0.1580	0.1590	0.1620	0.1580	0.1610	0.1540	0.1550	0.1550
29	0.1430	0.1430	0.1450	0.1430	0.1430	0.1430	0.1440	0.1470	0.1430	0.1480	0.1450	0.1450	0.1480	0.1460	0.1510	0.1440	0.1520	0.1410	0.1430	0.1430
30	0.1400	0.1420	0.1410	0.1390	0.1390	0.1380	0.1410	0.1420	0.1410	0.1430	0.1430	0.1400	0.1420	0.1420	0.1460	0.1400	0.1440	0.1380	0.1400	0.1400
31	0.1460	0.1490	0.1480	0.1520	0.1410	0.1490	0.1430	0.1530	0.1460	0.1480	0.1510	0.1510	0.1490	0.1510	0.1520	0.1470	0.1510	0.1420	0.1460	0.1470
32	0.1880	0.1890	0.1890	0.1880	0.1860	0.1880	0.1890	0.1900	0.1880	0.1890	0.1900	0.1890	0.1900	0.1910	0.1940	0.1890	0.1930	0.1850	0.1890	0.1880
33	0.0970	0.0940	0.0960	0.0930	0.0940	0.0940	0.0970	0.0990	0.0950	0.0960	0.0990	0.0960	0.1000	0.0960	0.1030	0.0950	0.1000	0.0890	0.0950	0.0940
34	0.0870	0.0860	0.0850	0.0830	0.0850	0.0780	0.0840	0.0890	0.0850	0.0880	0.0820	0.0870	0.0860	0.0830	0.0910	0.0860	0.0840	0.0850	0.0840	0.0810
35	0.1390	0.1380	0.1380	0.1380	0.1350	0.1350	0.1380	0.1400	0.1380	0.1380	0.1380	0.1360	0.1370	0.1360	0.1440	0.1360	0.1410	0.1330	0.1360	0.1360
36	0.1380	0.1370	0.1370	0.1360	0.1300	0.1350	0.1320	0.1390	0.1330	0.1330	0.1370	0.1370	0.1380	0.1340	0.1410	0.1380	0.1380	0.1300	0.1350	0.1340
37	0.1590	0.1610	0.1600	0.1580	0.1570	0.1570	0.1600	0.1620	0.1600	0.1610	0.1620	0.1590	0.1610	0.1620	0.1650	0.1590	0.1640	0.1570	0.1600	0.1580
38	0.1620	0.1660	0.1620	0.1620	0.1610	0.1610	0.1630	0.1670	0.1630	0.1650	0.1630	0.1630	0.1660	0.1660	0.1680	0.1640	0.1690	0.1600	0.1620	0.1620

Table C.5: Changes in NB TPR between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-0.0179	-0.0197	-0.0179	-0.0179	-0.0161	-0.0161	-0.0179	-0.0161	-0.0161	-0.0179	-0.0179	-0.0161	-0.0161	-0.0233	-0.0135	-0.0179	-0.0161	-0.0215	-0.0179	-0.0143
3	-0.0102	-0.0102	-0.0085	-0.0085	-0.0085	-0.0102	-0.0085	-0.0102	-0.0102	-0.0085	-0.0068	-0.0085	-0.0085	-0.0102	-0.0042	-0.0102	-0.0068	-0.0085	-0.0102	-0.0068
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
5	-0.0317	-0.0291	-0.0291	-0.0291	-0.0317	-0.0265	-0.0238	-0.0265	-0.0291	-0.0291	-0.0291	-0.0317	-0.0265	-0.0265	-0.0291	-0.0291	-0.0265	-0.0265	-0.0212	-0.0265
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	-0.0011	-0.0011	-0.0011	0.0000	-0.0011	0.0000	-0.0011	0.0000	-0.0011	0.0000	-0.0011	0.0000	0.0000	0.0015	-0.0023	-0.0011	-0.0011	0.0000	0.0000
8	-0.2159	-0.2292	-0.2392	-0.2259	-0.2392	-0.2193	-0.2259	-0.1993	-0.2259	-0.2093	-0.2292	-0.2126	-0.2193	-0.2292	-0.2200	-0.2193	-0.2093	-0.2193	-0.2027	-0.2126
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
10	-0.5346	-0.5392	-0.5392	-0.5346	-0.5346	-0.5346	-0.5484	-0.5207	-0.5438	-0.5161	-0.5668	-0.5300	-0.5484	-0.5484	-0.5273	-0.5622	-0.5853	-0.5438	-0.5346	-0.5207
11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
12	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-0.9974	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
14	-0.0989	-0.0856	-0.1043	-0.0936	-0.0963	-0.0989	-0.0936	-0.0963	-0.1016	-0.0936	-0.1016	-0.0882	-0.0909	-0.0856	-0.0910	-0.0936	-0.0963	-0.0936	-0.1016	-0.0882
15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
16	-0.1949	-0.1972	-0.2065	-0.1949	-0.1972	-0.1926	-0.1926	-0.1926	-0.1972	-0.1972	-0.1949	-0.1949	-0.1972	-0.1949	-0.1877	-0.1879	-0.1972	-0.1949	-0.1995	-0.1856
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
18	-0.1440	-0.1523	-0.1358	-0.1440	-0.1358	-0.1523	-0.1523	-0.1481	-0.1564	-0.1564	-0.1564	-0.1440	-0.1481	-0.1646	-0.1497	-0.1399	-0.1481	-0.1523	-0.1523	-0.1523
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
20	-0.2283	-0.2362	-0.2205	-0.2283	-0.2283	-0.2205	-0.2205	-0.2205	-0.2283	-0.2362	-0.2283	-0.2362	-0.2520	-0.2283	-0.2257	-0.2205	-0.2441	-0.2205	-0.2362	-0.2362
21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
22	-0.0522	-0.0448	-0.0522	-0.0398	-0.0547	-0.0448	-0.0323	-0.0522	-0.0448	-0.0423	-0.0498	-0.0547	-0.0522	-0.0522	-0.0471	-0.0522	-0.0448	-0.0448	-0.0448	-0.0547
23	-0.5837	-0.5773	-0.5923	-0.5794	-0.5773	-0.5837	-0.5837	-0.5773	-0.5858	-0.5687	-0.5794	-0.5944	-0.5901	-0.5966	-0.5768	-0.5880	-0.5966	-0.5901	-0.5944	-0.5901
24	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
25	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
26	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-0.9974	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
27	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
28	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
31	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
32	0.0000	0.0000	-0.0055	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
33	0.0000	-0.0013	0.0000	0.0000	0.0000	0.0000	-0.0013	-0.0013	0.0000	-0.0013	0.0000	-0.0013	-0.0013	-0.0013	0.0013	0.0000	-0.0013	-0.0013	-0.0013	0.0000
34	-0.0549	-0.0549	-0.0549	-0.0527	-0.0506	-0.0549	-0.0527	-0.0549	-0.0527	-0.0506	-0.0549	-0.0549	-0.0527	-0.0527	-0.0522	-0.0527	-0.0570	-0.0549	-0.0506	-0.0527
35	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
36	-0.1250	-0.1292	-0.1333	-0.1250	-0.1333	-0.1208	-0.1417	-0.1250	-0.1375	-0.1375	-0.1333	-0.1250	-0.1250	-0.1417	-0.1432	-0.1292	-0.1250	-0.1375	-0.1375	-0.1292
37	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000
38	-0.1513	-0.1447	-0.1513	-0.1447	-0.1513	-0.1447	-0.1513	-0.1447	-0.1447	-0.1447	-0.1579	-0.1513	-0.1447	-0.1513	-0.1553	-0.1447	-0.1447	-0.1447	-0.1579	-0.1447

Table C.6: Changes in NB FPR between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-0.4051	-0.4004	-0.3981	-0.3940	-0.3897	-0.3875	-0.4004	-0.4024	-0.4006	-0.4027	-0.4115	-0.3961	-0.4115	-0.4070	-0.4156	-0.4006	-0.4178	-0.3831	-0.3938	-0.3940
3	-0.6498	-0.6663	-0.6517	-0.6504	-0.6461	-0.6436	-0.6566	-0.6584	-0.6498	-0.6590	-0.6511	-0.6547	-0.6535	-0.6584	-0.6622	-0.6400	-0.6695	-0.6222	-0.6498	-0.6406
4	-0.2747	-0.2790	-0.3039	-0.2561	-0.2767	-0.2767	-0.2719	-0.2904	-0.2747	-0.2897	-0.2869	-0.2925	-0.2869	-0.2833	-0.2990	-0.2782	-0.3097	-0.2760	-0.2711	-0.2917
5	-0.7601	-0.7644	-0.7660	-0.7574	-0.7483	-0.7515	-0.7596	-0.7704	-0.7585	-0.7628	-0.7752	-0.7601	-0.7688	-0.7671	-0.7758	-0.7569	-0.7822	-0.7536	-0.7617	-0.7542
6	-0.0445	-0.0488	-0.0422	-0.0599	-0.0424	-0.0687	-0.0488	-0.0860	-0.0576	-0.0619	-0.0333	-0.0313	-0.0597	-0.0531	-0.0814	-0.0050	-0.0749	-0.0381	-0.0839	-0.0665
7	-0.0374	-0.0496	-0.0496	-0.0488	-0.0602	-0.0445	-0.0574	-0.0696	-0.0689	-0.0653	-0.0617	-0.0767	-0.0460	-0.0696	-0.0940	-0.0531	-0.0782	-0.0481	-0.0531	-0.0646
8	-0.8205	-0.8234	-0.8262	-0.8176	-0.8147	-0.8119	-0.8277	-0.8277	-0.8191	-0.8277	-0.8291	-0.8205	-0.8277	-0.8277	-0.8420	-0.8219	-0.8406	-0.8090	-0.8233	-0.8162
9	0.1595	0.1552	0.1552	0.1638	0.1681	0.1681	0.1552	0.1509	0.1595	0.1552	0.1509	0.1595	0.1509	0.1509	0.1422	0.1595	0.1422	0.1724	0.1595	0.1638
10	-0.8316	-0.8372	-0.8359	-0.8273	-0.8230	-0.8242	-0.8384	-0.8428	-0.8316	-0.8359	-0.8415	-0.8341	-0.8440	-0.8415	-0.8501	-0.8341	-0.8488	-0.8186	-0.8341	-0.8298
11	0.1595	0.1552	0.1552	0.1638	0.1681	0.1681	0.1552	0.1509	0.1595	0.1552	0.1509	0.1595	0.1509	0.1509	0.1422	0.1595	0.1422	0.1724	0.1595	0.1638
12	-0.8405	-0.8448	-0.8448	-0.8362	-0.8319	-0.8319	-0.8448	-0.8491	-0.8405	-0.8448	-0.8491	-0.8405	-0.8491	-0.8491	-0.8578	-0.8405	-0.8578	-0.8276	-0.8405	-0.8362
13	0.1595	0.1552	0.1552	0.1638	0.1681	0.1681	0.1552	0.1509	0.1595	0.1552	0.1509	0.1595	0.1509	0.1509	0.1422	0.1595	0.1422	0.1724	0.1595	0.1638
14	-0.6664	-0.6707	-0.6659	-0.6653	-0.6578	-0.6578	-0.6691	-0.6750	-0.6696	-0.6707	-0.6766	-0.6680	-0.6686	-0.6766	-0.6820	-0.6648	-0.6836	-0.6535	-0.6664	-0.6637
15	-0.1560	-0.1670	-0.1804	-0.1852	-0.1675	-0.1608	-0.1670	-0.1713	-0.1627	-0.1603	-0.1914	-0.1761	-0.1847	-0.1780	-0.1732	-0.1761	-0.1866	-0.1497	-0.1425	-0.1785
16	-0.7157	-0.7236	-0.7236	-0.7097	-0.7106	-0.7071	-0.7253	-0.7208	-0.7193	-0.7271	-0.7226	-0.7193	-0.7208	-0.7314	-0.7365	-0.7140	-0.7400	-0.7028	-0.7193	-0.7132
17	-0.1108	-0.1016	-0.1151	-0.0389	-0.1022	-0.0616	-0.1016	-0.0654	-0.0973	-0.0475	-0.0924	-0.1108	-0.1059	-0.1059	-0.0875	-0.0973	-0.0875	-0.0843	-0.0702	-0.0659
18	-0.7996	-0.8065	-0.8039	-0.7966	-0.7909	-0.7909	-0.8052	-0.8095	-0.8009	-0.8065	-0.8108	-0.8035	-0.8095	-0.8108	-0.8181	-0.8009	-0.8155	-0.7866	-0.8009	-0.7966
19	0.1595	0.1295	0.1295	0.1382	0.1681	0.1681	0.1552	0.1509	0.1595	0.1552	0.1509	0.1595	0.1252	0.1509	0.1422	0.1082	0.1422	0.1724	0.1595	0.1638
20	-0.8176	-0.8219	-0.8231	-0.8144	-0.8101	-0.8090	-0.8208	-0.8274	-0.8210	-0.8242	-0.8262	-0.8188	-0.8239	-0.8274	-0.8383	-0.8176	-0.8348	-0.8047	-0.8176	-0.8133
21	0.1595	0.1552	0.1552	0.1638	0.1681	0.1681	0.1552	0.1509	0.1595	0.1552	0.1509	0.1595	0.1509	0.1509	0.1422	0.1595	0.1422	0.1724	0.1595	0.1638
22	-0.6482	-0.6492	-0.6508	-0.6406	-0.6413	-0.6396	-0.6508	-0.6535	-0.6499	-0.6542	-0.6552	-0.6499	-0.6585	-0.6585	-0.6638	-0.6465	-0.6655	-0.6319	-0.6465	-0.6406
23	-0.7787	-0.7849	-0.7849	-0.7763	-0.7720	-0.7776	-0.7886	-0.7930	-0.7806	-0.7830	-0.7930	-0.7806	-0.7911	-0.7930	-0.8016	-0.7806	-0.7997	-0.7714	-0.7881	-0.7782
24	0.1537	0.1494	0.1552	0.1522	0.1623	0.1623	0.1494	0.1451	0.1421	0.1436	0.1393	0.1537	0.1509	0.1451	0.1307	0.1479	0.1365	0.1609	0.1537	0.1580
25	0.1595	0.1552	0.1552	0.1638	0.1681	0.1681	0.1552	0.1509	0.1595	0.1552	0.1509	0.1595	0.1509	0.1509	0.1422	0.1595	0.1422	0.1724	0.1595	0.1638
26	-0.8405	-0.8448	-0.8448	-0.8362	-0.8319	-0.8319	-0.8448	-0.8491	-0.8405	-0.8448	-0.8491	-0.8405	-0.8491	-0.8491	-0.8578	-0.8405	-0.8578	-0.8276	-0.8405	-0.8362
27	-0.6303	-0.6192	-0.6346	-0.6106	-0.6165	-0.6063	-0.6294	-0.6235	-0.6149	-0.6294	-0.5927	-0.6251	-0.6338	-0.6081	-0.6372	-0.6046	-0.6321	-0.6173	-0.5944	-0.6106
28	-0.7609	-0.7652	-0.7631	-0.7544	-0.7543	-0.7459	-0.7673	-0.7674	-0.7609	-0.7631	-0.7674	-0.7609	-0.7674	-0.7695	-0.7760	-0.7629	-0.7781	-0.7479	-0.7567	-0.7544
29	-0.6795	-0.6807	-0.6869	-0.6783	-0.6771	-0.6771	-0.6838	-0.6943	-0.6795	-0.6962	-0.6881	-0.6857	-0.6974	-0.6912	-0.7030	-0.6826	-0.7122	-0.6697	-0.6795	-0.6783
30	-0.6708	-0.6782	-0.6751	-0.6665	-0.6652	-0.6621	-0.6751	-0.6794	-0.6739	-0.6812	-0.6825	-0.6708	-0.6794	-0.6794	-0.6880	-0.6708	-0.6880	-0.6609	-0.6708	-0.6695
31	-0.0372	-0.0743	-0.0579	-0.1477	0.0206	-0.1106	0.0240	-0.1278	-0.0372	-0.0579	-0.0950	-0.1192	-0.0623	-0.0950	-0.0545	-0.0536	-0.0709	-0.0079	-0.0372	-0.0657
32	-0.8320	-0.8363	-0.8375	-0.8289	-0.8233	-0.8258	-0.8363	-0.8406	-0.8320	-0.8363	-0.8406	-0.8332	-0.8406	-0.8418	-0.8492	-0.8332	-0.8504	-0.8190	-0.8332	-0.8289
33	-0.3694	-0.3593	-0.3641	-0.3506	-0.3559	-0.3559	-0.3737	-0.3828	-0.3597	-0.3689	-0.3780	-0.3694	-0.3876	-0.3684	-0.3914	-0.3597	-0.3866	-0.3372	-0.3646	-0.3554
34	-0.6846	-0.6851	-0.6832	-0.6727	-0.6722	-0.6627	-0.6794	-0.6932	-0.6789	-0.6851	-0.6799	-0.6846	-0.6856	-0.6799	-0.6981	-0.6808	-0.6905	-0.6736	-0.6751	-0.6689
35	-0.0734	-0.0503	-0.0503	-0.0691	-0.0374	-0.0374	-0.0503	-0.0683	-0.0597	-0.0503	-0.0409	-0.0323	-0.0272	-0.0135	-0.0769	-0.0323	-0.0632	-0.0194	-0.0323	-0.0417
36	-0.8050	-0.8080	-0.8093	-0.7994	-0.7911	-0.7937	-0.8054	-0.8123	-0.8024	-0.8054	-0.8123	-0.8037	-0.8110	-0.8110	-0.8249	-0.8063	-0.8170	-0.7894	-0.8050	-0.7980
37	0.1595	0.1274	0.1552	0.1638	0.1681	0.1681	0.1552	0.1231	0.1317	0.1274	0.1231	0.1595	0.1509	0.1231	0.1422	0.1595	0.1145	0.1446	0.1317	0.1638
38	-0.8287	-0.8354	-0.8319	-0.8244	-0.8213	-0.8201	-0.8330	-0.8397	-0.8287	-0.8342	-0.8373	-0.8299	-0.8385	-0.8397	-0.8471	-0.8299	-0.8483	-0.8158	-0.8299	-0.8244

Table C.7: Changes in QDA accuracy between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-0.0030	-0.0040	-0.0050	-0.0050	-0.0060	-0.0050	0.0010	-0.0020	-0.0020	-0.0070	-0.0010	-0.0060	-0.0050	-0.0010	-0.0050	-0.0030	-0.0150	-0.0070	-0.0070	-0.0100
3	-0.0140	-0.0090	-0.0070	-0.0100	-0.0100	-0.0060	-0.0130	-0.0090	-0.0120	-0.0090	-0.0070	-0.0130	-0.0060	-0.0100	-0.0050	-0.0130	-0.0110	-0.0080	-0.0050	-0.0080
4	0.0040	0.0000	-0.0020	0.0020	0.0010	0.0010	0.0020	0.0000	-0.0020	0.0010	0.0000	0.0010	0.0000	0.0020	0.0020	-0.0020	-0.0020	0.0030	0.0000	0.0020
5	0.0120	0.0060	0.0120	0.0080	0.0090	0.0090	0.0020	0.0070	0.0130	0.0060	0.0070	0.0120	0.0120	0.0090	0.0120	0.0050	0.0070	0.0130	-0.0010	0.0100
6	0.0810	0.0810	0.0810	0.0810	0.0800	0.0810	0.0810	0.0810	0.0810	0.0800	0.0810	0.0810	0.0810	0.0800	0.0820	0.0810	0.0810	0.0810	0.0810	0.0810
7	0.1050	0.1050	0.1050	0.1050	0.1050	0.1050	0.1050	0.1050	0.1050	0.1050	0.1050	0.1050	0.1050	0.1050	0.1060	0.1050	0.1050	0.1050	0.1050	0.1050
8	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690	-0.0680	-0.0690	-0.0690	-0.0690	-0.0690	-0.0690
9	0.1680	0.1680	0.1680	0.1680	0.1680	0.1680	0.1680	0.1680	0.1680	0.1680	0.1680	0.1680	0.1680	0.1680	0.1690	0.1680	0.1680	0.1680	0.1680	0.1680
10	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0150	0.0160	0.0150	0.0150	0.0150	0.0150	0.0150
11	0.2120	0.2120	0.2120	0.2120	0.2120	0.2120	0.2120	0.2120	0.2120	0.2120	0.2120	0.2120	0.2120	0.2120	0.2130	0.2120	0.2120	0.2120	0.2120	0.2120
12	0.1810	0.1810	0.1810	0.1810	0.1810	0.1810	0.1810	0.1810	0.1810	0.1810	0.1810	0.1810	0.1810	0.1810	0.1820	0.1810	0.1810	0.1810	0.1810	0.1810
13	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900	0.0900	0.0910	0.0900	0.0900	0.0900	0.0900	0.0900
14	-0.0620	-0.0630	-0.0640	-0.0630	-0.0600	-0.0620	-0.0650	-0.0630	-0.0650	-0.0670	-0.0670	-0.0610	-0.0710	-0.0690	-0.0660	-0.0630	-0.0670	-0.0640	-0.0660	-0.0620
15	0.0830	0.0830	0.0830	0.0830	0.0830	0.0830	0.0830	0.0830	0.0830	0.0830	0.0830	0.0830	0.0830	0.0830	0.0840	0.0830	0.0830	0.0830	0.0830	0.0830
16	0.0620	0.0590	0.0640	0.0630	0.0610	0.0610	0.0620	0.0630	0.0660	0.0610	0.0630	0.0620	0.0620	0.0630	0.0640	0.0590	0.0670	0.0600	0.0610	0.0580
17	0.1580	0.1580	0.1580	0.1580	0.1580	0.1580	0.1580	0.1580	0.1580	0.1580	0.1580	0.1580	0.1580	0.1580	0.1590	0.1580	0.1580	0.1580	0.1580	0.1580
18	0.1280	0.1290	0.1250	0.1250	0.1270	0.1290	0.1230	0.1230	0.1230	0.1270	0.1230	0.1270	0.1270	0.1250	0.1280	0.1260	0.1250	0.1240	0.1250	0.1260
19	0.1930	0.1930	0.1930	0.1930	0.1930	0.1930	0.1930	0.1930	0.1930	0.1930	0.1930	0.1930	0.1930	0.1930	0.1940	0.1930	0.1930	0.1930	0.1930	0.1930
20	0.1630	0.1620	0.1640	0.1600	0.1610	0.1610	0.1610	0.1620	0.1640	0.1610	0.1630	0.1610	0.1610	0.1630	0.1650	0.1630	0.1620	0.1600	0.1610	0.1630
21	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0920	0.0910	0.0910	0.0910	0.0910	0.0910
22	-0.0300	-0.0310	-0.0250	-0.0320	-0.0260	-0.0260	-0.0270	-0.0210	-0.0260	-0.0260	-0.0270	-0.0330	-0.0260	-0.0290	-0.0280	-0.0230	-0.0250	-0.0260	-0.0270	-0.0210
23	0.0150	0.0110	0.0070	0.0180	0.0170	0.0110	0.0010	0.0100	-0.0090	0.0210	0.0080	0.0050	0.0040	0.0130	0.0170	0.0040	0.0040	0.0060	0.0080	0.0070
24	0.0590	0.0590	0.0590	0.0590	0.0590	0.0590	0.0590	0.0590	0.0600	0.0590	0.0600	0.0590	0.0590	0.0600	0.0600	0.0600	0.0590	0.0600	0.0590	0.0600
25	0.1650	0.1650	0.1650	0.1650	0.1650	0.1650	0.1650	0.1650	0.1650	0.1650	0.1650	0.1650	0.1650	0.1650	0.1660	0.1650	0.1650	0.1650	0.1650	0.1650
26	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340	-0.0330	-0.0340	-0.0340	-0.0340	-0.0340	-0.0340
27	0.0380	0.0400	0.0380	0.0390	0.0380	0.0380	0.0390	0.0390	0.0390	0.0420	0.0410	0.0390	0.0380	0.0390	0.0380	0.0400	0.0420	0.0370	0.0390	0.0410
28	0.2170	0.2170	0.2190	0.2190	0.2150	0.2190	0.2190	0.2180	0.2190	0.2180	0.2200	0.2150	0.2210	0.2160	0.2220	0.2220	0.2160	0.2190	0.2140	0.2190
29	0.0540	0.0540	0.0530	0.0560	0.0530	0.0510	0.0520	0.0530	0.0540	0.0530	0.0510	0.0530	0.0520	0.0510	0.0510	0.0520	0.0540	0.0540	0.0510	0.0540
30	0.1650	0.1670	0.1670	0.1640	0.1660	0.1680	0.1670	0.1680	0.1660	0.1660	0.1680	0.1670	0.1640	0.1640	0.1660	0.1640	0.1640	0.1670	0.1650	0.1640
31	0.1710	0.1710	0.1710	0.1710	0.1710	0.1710	0.1710	0.1710	0.1710	0.1710	0.1710	0.1710	0.1710	0.1710	0.1720	0.1710	0.1710	0.1710	0.1710	0.1710
32	0.1830	0.1850	0.1820	0.1830	0.1850	0.1830	0.1850	0.1840	0.1850	0.1830	0.1850	0.1820	0.1850	0.1860	0.1880	0.1850	0.1850	0.1830	0.1860	0.1860
33	0.0240	0.0240	0.0240	0.0240	0.0240	0.0240	0.0240	0.0240	0.0250	0.0240	0.0240	0.0240	0.0240	0.0240	0.0250	0.0240	0.0240	0.0240	0.0240	0.0240
34	0.0880	0.0910	0.0900	0.0850	0.0890	0.0860	0.0860	0.0890	0.0890	0.0930	0.0950	0.0910	0.0850	0.0900	0.0890	0.0870	0.0830	0.0860	0.0910	0.0820
35	0.1590	0.1590	0.1590	0.1590	0.1590	0.1590	0.1590	0.1590	0.1590	0.1590	0.1590	0.1590	0.1590	0.1590	0.1600	0.1590	0.1590	0.1590	0.1590	0.1590
36	0.1300	0.1290	0.1280	0.1280	0.1280	0.1240	0.1300	0.1270	0.1270	0.1290	0.1270	0.1280	0.1280	0.1280	0.1320	0.1260	0.1270	0.1260	0.1260	0.1290
37	0.1960	0.1960	0.1960	0.1960	0.1960	0.1960	0.1960	0.1960	0.1960	0.1960	0.1960	0.1960	0.1960	0.1960	0.1970	0.1960	0.1960	0.1960	0.1960	0.1960
38	0.1480	0.1490	0.1500	0.1500	0.1490	0.1480	0.1470	0.1490	0.1500	0.1470	0.1470	0.1470	0.1480	0.1490	0.1490	0.1500	0.1450	0.1480	0.1500	0.1470

Table C.8: Changes in QDA TPR between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-0.1521	-0.1467	-0.1485	-0.1521	-0.1503	-0.1503	-0.1485	-0.1485	-0.1431	-0.1521	-0.1467	-0.1521	-0.1521	-0.1431	-0.1525	-0.1467	-0.1628	-0.1574	-0.1521	-0.1574
3	-0.0355	-0.0338	-0.0355	-0.0406	-0.0372	-0.0338	-0.0372	-0.0372	-0.0389	-0.0355	-0.0355	-0.0406	-0.0321	-0.0406	-0.0325	-0.0355	-0.0355	-0.0355	-0.0338	-0.0355
4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
5	-0.5450	-0.5582	-0.5450	-0.5503	-0.5556	-0.5503	-0.5608	-0.5529	-0.5476	-0.5556	-0.5529	-0.5503	-0.5450	-0.5503	-0.5516	-0.5582	-0.5556	-0.5450	-0.5661	-0.5529
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
8	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-0.9987	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
10	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-0.9987	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
12	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-0.9987	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
14	-0.5909	-0.5882	-0.5989	-0.5936	-0.5882	-0.5882	-0.5989	-0.5936	-0.5989	-0.6070	-0.6043	-0.5856	-0.6123	-0.6070	-0.6003	-0.5936	-0.6043	-0.5963	-0.5989	-0.5882
15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
16	-0.2877	-0.2877	-0.2831	-0.2854	-0.2877	-0.2900	-0.2877	-0.2854	-0.2807	-0.2923	-0.2923	-0.2877	-0.2854	-0.2831	-0.2864	-0.2900	-0.2831	-0.2923	-0.2831	-0.2947
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
18	-0.3868	-0.3827	-0.3992	-0.3951	-0.3909	-0.3827	-0.4033	-0.4074	-0.3992	-0.3951	-0.4115	-0.3951	-0.3951	-0.4074	-0.3896	-0.3909	-0.3951	-0.3951	-0.3992	-0.3909
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
20	-0.5433	-0.5512	-0.5354	-0.5669	-0.5591	-0.5591	-0.5591	-0.5512	-0.5354	-0.5591	-0.5433	-0.5591	-0.5591	-0.5433	-0.5341	-0.5433	-0.5512	-0.5669	-0.5591	-0.5433
21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
22	-0.4751	-0.4776	-0.4627	-0.4801	-0.4652	-0.4652	-0.4677	-0.4527	-0.4652	-0.4652	-0.4677	-0.4826	-0.4652	-0.4726	-0.4713	-0.4577	-0.4627	-0.4652	-0.4677	-0.4527
23	-0.3326	-0.3476	-0.3498	-0.3498	-0.3391	-0.3412	-0.3627	-0.3562	-0.3798	-0.3326	-0.3519	-0.3519	-0.3519	-0.3455	-0.3463	-0.3627	-0.3584	-0.3498	-0.3455	-0.3455
24	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
25	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
26	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-0.9987	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
27	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
28	-0.0134	-0.0153	-0.0153	-0.0153	-0.0172	-0.0172	-0.0115	-0.0134	-0.0134	-0.0153	-0.0172	-0.0210	-0.0115	-0.0210	-0.0121	-0.0153	-0.0153	-0.0134	-0.0172	-0.0172
29	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
30	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
31	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
32	-0.2692	-0.2582	-0.2747	-0.2692	-0.2582	-0.2692	-0.2582	-0.2637	-0.2582	-0.2692	-0.2582	-0.2747	-0.2582	-0.2527	-0.2460	-0.2582	-0.2582	-0.2692	-0.2527	-0.2527
33	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
34	-0.2046	-0.2046	-0.2046	-0.2110	-0.2025	-0.2068	-0.2131	-0.2046	-0.2004	-0.1962	-0.1920	-0.1983	-0.2110	-0.2025	-0.2033	-0.2089	-0.2110	-0.2110	-0.2004	-0.2131
35	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
36	-0.4250	-0.4292	-0.4333	-0.4333	-0.4333	-0.4500	-0.4250	-0.4375	-0.4375	-0.4292	-0.4375	-0.4333	-0.4333	-0.4333	-0.4195	-0.4417	-0.4375	-0.4417	-0.4417	-0.4292
37	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000
38	-0.5526	-0.5461	-0.5395	-0.5395	-0.5461	-0.5526	-0.5592	-0.5461	-0.5395	-0.5592	-0.5592	-0.5592	-0.5526	-0.5461	-0.5513	-0.5395	-0.5724	-0.5526	-0.5395	-0.5592

Table C.9: Changes in QDA FPR between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-0.6599	-0.6508	-0.6508	-0.6553	-0.6508	-0.6531	-0.6644	-0.6576	-0.6508	-0.6508	-0.6576	-0.6531	-0.6553	-0.6531	-0.6553	-0.6531	-0.6463	-0.6576	-0.6508	-0.6508
3	-0.4499	-0.4597	-0.4670	-0.4670	-0.4621	-0.4670	-0.4548	-0.4645	-0.4597	-0.4621	-0.4670	-0.4597	-0.4645	-0.4670	-0.4670	-0.4523	-0.4572	-0.4645	-0.4694	-0.4645
4	-0.1886	-0.1744	-0.1673	-0.1815	-0.1779	-0.1779	-0.1815	-0.1744	-0.1673	-0.1779	-0.1744	-0.1779	-0.1744	-0.1815	-0.1779	-0.1673	-0.1673	-0.1851	-0.1744	-0.1815
5	-0.9775	-0.9759	-0.9775	-0.9743	-0.9791	-0.9759	-0.9711	-0.9743	-0.9807	-0.9743	-0.9743	-0.9807	-0.9775	-0.9759	-0.9807	-0.9743	-0.9759	-0.9791	-0.9695	-0.9791
6	-0.0066	-0.0066	-0.0066	-0.0066	0.0000	-0.0066	-0.0066	-0.0066	-0.0066	0.0000	-0.0066	-0.0066	-0.0066	0.0000	-0.0066	-0.0066	-0.0066	-0.0066	-0.0066	-0.0066
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
11	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	-0.8834	-0.8802	-0.8850	-0.8834	-0.8850	-0.8818	-0.8834	-0.8834	-0.8834	-0.8850	-0.8834	-0.8818	-0.8818	-0.8818	-0.8818	-0.8834	-0.8834	-0.8834	-0.8818	-0.8818
15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
16	-0.9192	-0.9139	-0.9192	-0.9192	-0.9174	-0.9192	-0.9192	-0.9192	-0.9209	-0.9209	-0.9244	-0.9192	-0.9174	-0.9174	-0.9209	-0.9156	-0.9244	-0.9192	-0.9139	-0.9174
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
18	-0.9868	-0.9868	-0.9868	-0.9855	-0.9868	-0.9868	-0.9855	-0.9868	-0.9841	-0.9881	-0.9881	-0.9881	-0.9881	-0.9894	-0.9868	-0.9855	-0.9855	-0.9841	-0.9868	-0.9855
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
22	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813	-0.8813
23	-0.8839	-0.8895	-0.8839	-0.9045	-0.8933	-0.8839	-0.8839	-0.8951	-0.8801	-0.8951	-0.8876	-0.8820	-0.8801	-0.8914	-0.8989	-0.8895	-0.8858	-0.8820	-0.8820	-0.8801
24	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0058	0.0000	-0.0058	0.0000	0.0000	-0.0058	0.0000	-0.0058	0.0000	-0.0058	0.0000	-0.0058
25	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
26	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
27	-0.0051	-0.0154	-0.0051	-0.0103	-0.0051	-0.0051	-0.0103	-0.0103	-0.0103	-0.0256	-0.0205	-0.0103	-0.0051	-0.0103	0.0000	-0.0154	-0.0256	0.0000	-0.0103	-0.0205
28	-0.9832	-0.9853	-0.9895	-0.9895	-0.9832	-0.9916	-0.9853	-0.9853	-0.9874	-0.9874	-0.9937	-0.9874	-0.9895	-0.9895	-0.9916	-0.9958	-0.9832	-0.9874	-0.9811	-0.9916
29	-0.4489	-0.4489	-0.4458	-0.4551	-0.4458	-0.4396	-0.4427	-0.4458	-0.4489	-0.4458	-0.4396	-0.4458	-0.4427	-0.4396	-0.4365	-0.4427	-0.4489	-0.4489	-0.4396	-0.4489
30	-0.7932	-0.7994	-0.7994	-0.7901	-0.7963	-0.8025	-0.7994	-0.8025	-0.7963	-0.7963	-0.8025	-0.7994	-0.7901	-0.7901	-0.7932	-0.7901	-0.7901	-0.7994	-0.7932	-0.7901
31	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
32	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
33	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0048	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
34	-0.9106	-0.9163	-0.9144	-0.9106	-0.9106	-0.9087	-0.9144	-0.9125	-0.9087	-0.9125	-0.9125	-0.9106	-0.9106	-0.9125	-0.9106	-0.9125	-0.9068	-0.9125	-0.9125	-0.9068
35	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
36	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000
37	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
38	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000

Table C.10: Changes in SVM accuracy between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0840	0.0880	0.0780	0.0860	0.0860	0.0820	0.0850	0.0830	0.0880	0.0880	0.0870	0.0910	0.0760	0.0890	0.0880	0.0820	0.0870	0.0800	0.0860	0.0980
3	0.1980	0.1960	0.1930	0.2050	0.2040	0.1970	0.1930	0.2030	0.1970	0.2020	0.2020	0.2030	0.1940	0.2050	0.2010	0.1970	0.1990	0.2020	0.2030	0.2050
4	0.2260	0.2270	0.2200	0.2270	0.2230	0.2260	0.2190	0.2280	0.2270	0.2210	0.2300	0.2300	0.2250	0.2270	0.2250	0.2170	0.2260	0.2230	0.2240	0.2290
5	0.2010	0.2000	0.1970	0.2000	0.2070	0.2030	0.1970	0.2010	0.2090	0.1980	0.2040	0.2000	0.1960	0.2030	0.2090	0.2010	0.2030	0.2010	0.2040	0.2060
6	0.2130	0.2140	0.2140	0.2180	0.2180	0.2150	0.2160	0.2170	0.2190	0.2170	0.2190	0.2230	0.2140	0.2170	0.2170	0.2140	0.2210	0.2110	0.2160	0.2240
7	-0.0310	-0.0300	-0.0370	-0.0290	-0.0300	-0.0230	-0.0280	-0.0300	-0.0230	-0.0200	-0.0260	-0.0260	-0.0290	-0.0260	-0.0220	-0.0270	-0.0190	-0.0210	-0.0220	-0.0190
8	0.0090	0.0110	0.0090	0.0160	0.0150	0.0090	0.0110	0.0120	0.0160	0.0130	0.0200	0.0170	0.0070	0.0110	0.0180	0.0100	0.0170	0.0130	0.0150	0.0200
9	-0.0450	-0.0470	-0.0490	-0.0400	-0.0480	-0.0370	-0.0360	-0.0450	-0.0340	-0.0450	-0.0400	-0.0410	-0.0450	-0.0380	-0.0340	-0.0430	-0.0360	-0.0400	-0.0380	-0.0340
10	-0.0110	-0.0130	-0.0120	-0.0040	-0.0120	-0.0070	-0.0160	-0.0100	-0.0100	-0.0140	-0.0080	-0.0030	-0.0120	-0.0140	-0.0040	-0.0120	-0.0120	-0.0080	-0.0090	-0.0040
11	-0.1010	-0.0950	-0.1100	-0.0930	-0.0920	-0.1070	-0.1010	-0.1040	-0.0750	-0.0900	-0.0990	-0.0720	-0.1070	-0.1040	-0.0870	-0.1020	-0.1090	-0.0930	-0.0920	-0.0960
12	0.0100	0.0080	0.0040	0.0180	0.0150	0.0180	0.0100	0.0180	0.0150	0.0090	0.0120	0.0150	0.0090	0.0170	0.0180	0.0130	0.0170	0.0130	0.0180	0.0230
13	0.0840	0.0890	0.0870	0.0950	0.0970	0.0900	0.0880	0.0900	0.0930	0.0920	0.0870	0.0970	0.0860	0.0880	0.0950	0.0910	0.0930	0.0950	0.0930	0.0990
14	0.1100	0.1160	0.1110	0.1180	0.1240	0.1160	0.1170	0.1200	0.1230	0.1180	0.1190	0.1220	0.1130	0.1220	0.1200	0.1190	0.1260	0.1170	0.1180	0.1270
15	0.1240	0.1250	0.1250	0.1330	0.1300	0.1350	0.1310	0.1300	0.1360	0.1230	0.1300	0.1370	0.1260	0.1350	0.1330	0.1330	0.1310	0.1290	0.1320	0.1360
16	0.1130	0.1120	0.1130	0.1210	0.1160	0.1140	0.1190	0.1170	0.1160	0.1110	0.1210	0.1180	0.1150	0.1120	0.1240	0.1120	0.1210	0.1170	0.1190	0.1190
17	0.1720	0.1730	0.1690	0.1860	0.1810	0.1780	0.1740	0.1740	0.1800	0.1810	0.1790	0.1820	0.1750	0.1790	0.1820	0.1730	0.1820	0.1770	0.1760	0.1900
18	0.1850	0.1820	0.1780	0.1870	0.1850	0.1830	0.1830	0.1820	0.1900	0.1790	0.1810	0.1880	0.1830	0.1860	0.1850	0.1850	0.1850	0.1850	0.1860	0.1910
19	0.1860	0.1890	0.1850	0.1930	0.1890	0.1860	0.1860	0.1910	0.1900	0.1840	0.1880	0.1960	0.1860	0.1910	0.1900	0.1860	0.1920	0.1880	0.1870	0.1970
20	0.1800	0.1790	0.1740	0.1840	0.1830	0.1810	0.1790	0.1820	0.1820	0.1810	0.1830	0.1870	0.1800	0.1840	0.1830	0.1850	0.1850	0.1800	0.1780	0.1910
21	0.0760	0.0760	0.0770	0.0850	0.0810	0.0820	0.0810	0.0820	0.0800	0.0770	0.0860	0.0900	0.0870	0.0850	0.0900	0.0820	0.0930	0.0820	0.0840	0.0960
22	0.1320	0.1320	0.1310	0.1380	0.1390	0.1340	0.1330	0.1360	0.1370	0.1330	0.1330	0.1430	0.1310	0.1340	0.1400	0.1330	0.1390	0.1360	0.1370	0.1430
23	0.2160	0.2190	0.2140	0.2200	0.2230	0.2150	0.2140	0.2190	0.2210	0.2180	0.2190	0.2260	0.2150	0.2210	0.2230	0.2180	0.2240	0.2150	0.2220	0.2300
24	0.2110	0.2120	0.2180	0.2210	0.2210	0.2180	0.2160	0.2160	0.2180	0.2170	0.2200	0.2270	0.2170	0.2220	0.2190	0.2140	0.2230	0.2180	0.2210	0.2220
25	-0.1050	-0.0990	-0.1090	-0.1000	-0.1040	-0.1070	-0.1020	-0.0990	-0.0980	-0.1080	-0.1010	-0.0970	-0.1020	-0.1060	-0.0970	-0.1050	-0.0990	-0.1010	-0.1020	-0.0960
26	-0.0100	-0.0090	-0.0140	-0.0030	-0.0050	-0.0050	-0.0120	-0.0030	-0.0030	-0.0050	-0.0010	-0.0020	-0.0060	-0.0040	-0.0010	-0.0070	-0.0010	-0.0050	-0.0080	0.0010
27	0.2070	0.2090	0.2020	0.2230	0.2240	0.2180	0.2090	0.2130	0.2150	0.2160	0.2170	0.2180	0.2110	0.2180	0.2180	0.2150	0.2190	0.2130	0.2200	0.2260
28	0.2360	0.2380	0.2330	0.2420	0.2380	0.2370	0.2340	0.2350	0.2410	0.2380	0.2410	0.2430	0.2360	0.2390	0.2420	0.2390	0.2440	0.2380	0.2390	0.2480
29	0.2470	0.2470	0.2460	0.2530	0.2530	0.2520	0.2470	0.2510	0.2570	0.2510	0.2530	0.2570	0.2480	0.2550	0.2550	0.2530	0.2570	0.2500	0.2520	0.2600
30	0.2460	0.2450	0.2380	0.2500	0.2500	0.2480	0.2410	0.2480	0.2510	0.2470	0.2510	0.2530	0.2430	0.2510	0.2510	0.2480	0.2510	0.2460	0.2500	0.2580
31	0.1320	0.1330	0.1250	0.1340	0.1350	0.1330	0.1310	0.1350	0.1410	0.1350	0.1350	0.1400	0.1370	0.1310	0.1360	0.1330	0.1390	0.1340	0.1380	0.1410
32	0.1320	0.1310	0.1280	0.1370	0.1350	0.1310	0.1290	0.1330	0.1370	0.1290	0.1330	0.1390	0.1330	0.1370	0.1390	0.1350	0.1380	0.1330	0.1360	0.1410
33	0.2000	0.2030	0.1940	0.2060	0.2020	0.2020	0.2070	0.2050	0.2120	0.2030	0.2060	0.2080	0.2030	0.2090	0.2130	0.1990	0.2080	0.2030	0.2070	0.2110
34	0.2070	0.2100	0.2080	0.2170	0.2180	0.2030	0.2050	0.2120	0.2170	0.2090	0.2150	0.2180	0.2080	0.2110	0.2130	0.2120	0.2160	0.2050	0.2130	0.2180
35	0.1910	0.1860	0.1870	0.1930	0.1910	0.1870	0.1870	0.1940	0.1940	0.1900	0.1910	0.1940	0.1890	0.1900	0.1990	0.1920	0.1930	0.1880	0.1920	0.2000
36	0.2040	0.2010	0.2000	0.2060	0.2110	0.2060	0.2020	0.2030	0.2060	0.2040	0.2030	0.2100	0.2000	0.2040	0.2050	0.2050	0.2080	0.2040	0.2050	0.2100
37	0.2150	0.2140	0.2100	0.2200	0.2190	0.2130	0.2110	0.2150	0.2240	0.2150	0.2180	0.2200	0.2150	0.2230	0.2190	0.2160	0.2230	0.2210	0.2190	0.2260
38	0.1940	0.1940	0.1890	0.2000	0.2000	0.1950	0.1960	0.1950	0.2000	0.1950	0.2020	0.2020	0.1940	0.2000	0.1980	0.1960	0.2010	0.1950	0.1950	0.2040

Table C.11: Changes in SVM TPR between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.2435	0.2414	0.2312	0.2475	0.2377	0.2294	0.2400	0.2390	0.2413	0.2466	0.2387	0.2457	0.2307	0.2400	0.2434	0.2338	0.2390	0.2320	0.2359	0.2475
3	0.2136	0.2080	0.2101	0.2226	0.2149	0.2050	0.2102	0.2196	0.2098	0.2200	0.2140	0.2175	0.2046	0.2237	0.2154	0.2059	0.2111	0.2144	0.2183	0.2141
4	0.2859	0.2806	0.2779	0.2856	0.2816	0.2793	0.2789	0.2871	0.2830	0.2816	0.2846	0.2856	0.2820	0.2817	0.2828	0.2708	0.2843	0.2777	0.2802	0.2815
5	0.2913	0.2874	0.2834	0.2873	0.2900	0.2834	0.2886	0.2939	0.2952	0.2873	0.2873	0.2846	0.2821	0.2886	0.2939	0.2887	0.2886	0.2860	0.2900	0.2873
6	0.2597	0.2546	0.2580	0.2624	0.2610	0.2533	0.2620	0.2623	0.2598	0.2633	0.2584	0.2648	0.2558	0.2585	0.2602	0.2559	0.2623	0.2524	0.2598	0.2624
7	-0.0251	-0.0290	-0.0280	-0.0246	-0.0272	-0.0211	-0.0217	-0.0225	-0.0215	-0.0123	-0.0253	-0.0246	-0.0267	-0.0240	-0.0176	-0.0254	-0.0168	-0.0163	-0.0181	-0.0201
8	0.3151	0.3112	0.3099	0.3190	0.3164	0.3099	0.3151	0.3177	0.3164	0.3164	0.3138	0.3190	0.3112	0.3151	0.3203	0.3125	0.3177	0.3125	0.3164	0.3190
9	-0.0161	-0.0232	-0.0224	-0.0122	-0.0223	-0.0138	-0.0065	-0.0156	-0.0094	-0.0159	-0.0163	-0.0165	-0.0200	-0.0118	-0.0066	-0.0187	-0.0114	-0.0155	-0.0116	-0.0111
10	0.3151	0.3112	0.3099	0.3190	0.3164	0.3099	0.3151	0.3177	0.3164	0.3164	0.3138	0.3190	0.3112	0.3151	0.3203	0.3125	0.3177	0.3125	0.3164	0.3190
11	-0.0594	-0.0561	-0.0707	-0.0524	-0.0520	-0.0728	-0.0594	-0.0629	-0.0336	-0.0458	-0.0627	-0.0279	-0.0694	-0.0645	-0.0450	-0.0651	-0.0721	-0.0538	-0.0530	-0.0616
12	0.1975	0.1936	0.1922	0.2014	0.1988	0.1922	0.1975	0.2001	0.1988	0.1988	0.1962	0.2014	0.1936	0.1975	0.2027	0.1949	0.2001	0.1949	0.1988	0.2014
13	0.1030	0.1037	0.1048	0.1139	0.1136	0.1024	0.1076	0.1091	0.1078	0.1124	0.1005	0.1127	0.1014	0.1030	0.1129	0.1062	0.1079	0.1109	0.1101	0.1116
14	0.3151	0.3112	0.3099	0.3190	0.3164	0.3099	0.3151	0.3177	0.3164	0.3164	0.3138	0.3190	0.3112	0.3151	0.3203	0.3125	0.3177	0.3125	0.3164	0.3190
15	0.1811	0.1749	0.1783	0.1898	0.1836	0.1830	0.1870	0.1849	0.1883	0.1789	0.1834	0.1886	0.1784	0.1894	0.1852	0.1856	0.1826	0.1797	0.1860	0.1839
16	0.1272	0.1209	0.1196	0.1334	0.1238	0.1196	0.1272	0.1251	0.1215	0.1238	0.1282	0.1264	0.1233	0.1202	0.1370	0.1199	0.1275	0.1199	0.1285	0.1195
17	0.2190	0.2140	0.2138	0.2315	0.2235	0.2170	0.2201	0.2205	0.2225	0.2279	0.2198	0.2240	0.2172	0.2212	0.2264	0.2153	0.2238	0.2196	0.2192	0.2294
18	0.2740	0.2618	0.2605	0.2737	0.2670	0.2605	0.2657	0.2683	0.2670	0.2711	0.2644	0.2696	0.2618	0.2657	0.2709	0.2672	0.2683	0.2631	0.2670	0.2779
19	0.2339	0.2321	0.2308	0.2399	0.2332	0.2266	0.2339	0.2386	0.2332	0.2321	0.2316	0.2399	0.2300	0.2350	0.2360	0.2293	0.2355	0.2313	0.2321	0.2378
20	0.2994	0.2954	0.2941	0.2954	0.3007	0.2941	0.2994	0.3020	0.3007	0.3007	0.2981	0.3033	0.2954	0.2915	0.3046	0.2968	0.2941	0.2968	0.2928	0.3033
21	0.0939	0.0888	0.0934	0.1025	0.0952	0.0934	0.0997	0.1000	0.0929	0.0952	0.0996	0.1048	0.1028	0.0997	0.1073	0.0960	0.1082	0.0960	0.0999	0.1083
22	0.3151	0.3112	0.3099	0.3190	0.3164	0.3099	0.3151	0.3177	0.3164	0.3164	0.3138	0.3190	0.3112	0.3151	0.3203	0.3125	0.3177	0.3125	0.3164	0.3190
23	0.2572	0.2533	0.2498	0.2611	0.2606	0.2455	0.2507	0.2576	0.2542	0.2606	0.2559	0.2632	0.2511	0.2572	0.2602	0.2524	0.2598	0.2503	0.2563	0.2654
24	0.2643	0.2616	0.2664	0.2706	0.2680	0.2603	0.2692	0.2681	0.2632	0.2705	0.2654	0.2731	0.2628	0.2679	0.2695	0.2617	0.2693	0.2629	0.2680	0.2670
25	-0.0825	-0.0811	-0.0877	-0.0786	-0.0855	-0.0920	-0.0793	-0.0767	-0.0791	-0.0855	-0.0838	-0.0786	-0.0832	-0.0879	-0.0763	-0.0873	-0.0810	-0.0830	-0.0823	-0.0808
26	0.3151	0.3112	0.3099	0.3190	0.3164	0.3099	0.3151	0.3177	0.3164	0.3164	0.3138	0.3190	0.3112	0.3151	0.3203	0.3125	0.3177	0.3125	0.3164	0.3190
27	0.2431	0.2404	0.2354	0.2594	0.2593	0.2478	0.2443	0.2494	0.2468	0.2543	0.2480	0.2494	0.2441	0.2505	0.2532	0.2467	0.2506	0.2442	0.2543	0.2557
28	0.3132	0.3093	0.3099	0.3190	0.3145	0.3099	0.3132	0.3158	0.3145	0.3145	0.3138	0.3171	0.3093	0.3132	0.3184	0.3125	0.3177	0.3106	0.3145	0.3171
29	0.2959	0.2920	0.2936	0.3028	0.3002	0.2951	0.2959	0.3000	0.3016	0.3002	0.2990	0.3028	0.2964	0.3003	0.3041	0.2977	0.3015	0.2963	0.2987	0.3028
30	0.2855	0.2801	0.2803	0.2894	0.2868	0.2833	0.2796	0.2896	0.2868	0.2868	0.2872	0.2894	0.2787	0.2914	0.2892	0.2844	0.2866	0.2829	0.2883	0.2924
31	0.1724	0.1685	0.1629	0.1731	0.1716	0.1661	0.1713	0.1750	0.1769	0.1758	0.1700	0.1763	0.1738	0.1671	0.1744	0.1687	0.1750	0.1698	0.1758	0.1742
32	0.3151	0.3112	0.3099	0.3190	0.3164	0.3099	0.3151	0.3177	0.3164	0.3164	0.3138	0.3190	0.3112	0.3151	0.3203	0.3125	0.3177	0.3125	0.3164	0.3190
33	0.2469	0.2455	0.2367	0.2508	0.2444	0.2417	0.2520	0.2533	0.2545	0.2495	0.2456	0.2508	0.2455	0.2520	0.2572	0.2431	0.2508	0.2443	0.2507	0.2470
34	0.2835	0.2880	0.2846	0.2937	0.2932	0.2761	0.2813	0.2945	0.2932	0.2848	0.2864	0.2916	0.2838	0.2813	0.2866	0.2851	0.2903	0.2787	0.2869	0.2916
35	0.2353	0.2249	0.2290	0.2370	0.2333	0.2258	0.2320	0.2390	0.2344	0.2355	0.2297	0.2349	0.2303	0.2320	0.2416	0.2316	0.2325	0.2284	0.2355	0.2370
36	0.2984	0.2904	0.2891	0.2940	0.2997	0.2932	0.2984	0.2927	0.2914	0.2997	0.2846	0.2982	0.2862	0.2943	0.2995	0.2958	0.2927	0.2917	0.2997	0.2940
37	0.2632	0.2573	0.2560	0.2671	0.2635	0.2539	0.2591	0.2627	0.2677	0.2635	0.2609	0.2640	0.2593	0.2674	0.2653	0.2596	0.2669	0.2648	0.2645	0.2671
38	0.3019	0.2915	0.2902	0.3059	0.3098	0.2967	0.3085	0.3046	0.3032	0.2967	0.2941	0.3124	0.2915	0.3019	0.3072	0.2993	0.3046	0.2993	0.2967	0.3059

Table C.12: Changes in SVM FPR between datasets over 20 repetitions.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.2185	0.1968	0.2120	0.2167	0.1990	0.1904	0.2117	0.2142	0.1970	0.2140	0.1947	0.1970	0.2124	0.1945	0.2056	0.2013	0.1970	0.2036	0.1988	0.1773
3	-0.0839	-0.0962	-0.0809	-0.0901	-0.1036	-0.1073	-0.0766	-0.0882	-0.0957	-0.0839	-0.1030	-0.0981	-0.0944	-0.0938	-0.0919	-0.1011	-0.0981	-0.1011	-0.0944	-0.1135
4	-0.0358	-0.0566	-0.0366	-0.0409	-0.0388	-0.0581	-0.0287	-0.0401	-0.0503	-0.0287	-0.0574	-0.0538	-0.0452	-0.0531	-0.0416	-0.0460	-0.0431	-0.0495	-0.0452	-0.0632
5	-0.0134	-0.0274	-0.0177	-0.0172	-0.0339	-0.0409	-0.0086	-0.0129	-0.0366	-0.0102	-0.0350	-0.0269	-0.0215	-0.0290	-0.0296	-0.0274	-0.0301	-0.0290	-0.0263	-0.0431
6	-0.0640	-0.0878	-0.0749	-0.0792	-0.0812	-0.0964	-0.0706	-0.0749	-0.0921	-0.0706	-0.0987	-0.0921	-0.0835	-0.0878	-0.0835	-0.0812	-0.0921	-0.0812	-0.0769	-0.1050
7	-0.1034	-0.1207	-0.0684	-0.1121	-0.1128	-0.1136	-0.1034	-0.0920	-0.1250	-0.1034	-0.1250	-0.1250	-0.1164	-0.1207	-0.1164	-0.1207	-0.1250	-0.1049	-0.1164	-0.1379
8	0.2642	0.2455	0.2556	0.2527	0.2455	0.2412	0.2614	0.2599	0.2412	0.2599	0.2327	0.2427	0.2542	0.2498	0.2484	0.2484	0.2412	0.2441	0.2484	0.2297
9	-0.0878	-0.1051	-0.0921	-0.0964	-0.0894	-0.0981	-0.0878	-0.0765	-0.1094	-0.0878	-0.1094	-0.1094	-0.1008	-0.0894	-0.1008	-0.1051	-0.1094	-0.1051	-0.1008	-0.1067
10	0.2503	0.2369	0.2435	0.2391	0.2407	0.2219	0.2567	0.2486	0.2352	0.2554	0.2300	0.2288	0.2387	0.2420	0.2374	0.2369	0.2390	0.2318	0.2399	0.2209
11	-0.1034	-0.0207	-0.1078	-0.1121	-0.0207	-0.1293	-0.1034	-0.1078	0.0750	-0.0034	-0.1250	0.1750	-0.1164	-0.0207	-0.0164	-0.1207	-0.1250	-0.0207	-0.1164	-0.1379
12	0.1600	0.1459	0.1588	0.1482	0.1427	0.1278	0.1600	0.1504	0.1395	0.1621	0.1405	0.1416	0.1481	0.1396	0.1460	0.1417	0.1384	0.1417	0.1428	0.1234
13	-0.1034	-0.1207	-0.1078	-0.1121	-0.1207	-0.1293	-0.1034	-0.1078	-0.1250	-0.1034	-0.1250	-0.1250	-0.1164	-0.1207	-0.1164	-0.1207	-0.1250	-0.1207	-0.1164	-0.1379
14	0.1458	0.1205	0.1351	0.1323	0.1141	0.1135	0.1346	0.1303	0.1130	0.1346	0.1162	0.1178	0.1280	0.1157	0.1280	0.1173	0.1098	0.1205	0.1264	0.1017
15	0.0845	0.0538	0.0667	0.0759	0.0672	0.0385	0.0710	0.0667	0.0562	0.0778	0.0696	0.0495	0.0648	0.0672	0.0514	0.0605	0.0562	0.0538	0.0648	0.0366
16	0.0231	0.0076	0.0117	0.0110	0.0058	-0.0028	0.0125	0.0135	0.0015	0.0248	-0.0037	0.0033	0.0066	0.0094	0.0066	0.0076	-0.0020	-0.0012	0.0066	-0.0114
17	-0.0359	-0.0666	-0.0402	-0.0580	-0.0666	-0.0753	-0.0494	-0.0402	-0.0574	-0.0494	-0.0709	-0.0709	-0.0623	-0.0666	-0.0623	-0.0531	-0.0709	-0.0531	-0.0623	-0.0839
18	-0.0097	-0.0243	-0.0113	-0.0156	-0.0229	-0.0329	-0.0097	-0.0087	-0.0325	-0.0017	-0.0233	-0.0272	-0.0226	-0.0256	-0.0160	-0.0256	-0.0246	-0.0269	-0.0213	-0.0375
19	-0.0522	-0.0694	-0.0565	-0.0608	-0.0694	-0.0780	-0.0522	-0.0565	-0.0737	-0.0522	-0.0481	-0.0737	-0.0651	-0.0694	-0.0651	-0.0694	-0.0737	-0.0694	-0.0651	-0.0866
20	-0.0072	-0.0222	-0.0081	-0.0158	-0.0222	-0.0319	-0.0061	-0.0104	-0.0242	-0.0072	-0.0276	-0.0276	-0.0202	-0.0256	-0.0156	-0.0279	-0.0276	-0.0222	-0.0144	-0.0417
21	-0.1034	-0.1207	-0.1078	-0.1121	-0.1207	-0.1293	-0.1034	-0.1078	-0.1250	-0.1034	-0.1250	-0.1250	-0.1164	-0.1207	-0.1164	-0.1207	-0.1250	-0.1207	-0.1164	-0.1379
22	0.1206	0.1051	0.1130	0.1103	0.1000	0.0948	0.1190	0.1146	0.1008	0.1206	0.1041	0.0941	0.1094	0.1067	0.1060	0.1051	0.0991	0.1000	0.1060	0.0861
23	-0.0604	-0.0814	-0.0684	-0.0671	-0.0795	-0.0862	-0.0622	-0.0666	-0.0838	-0.0604	-0.0801	-0.0838	-0.0733	-0.0795	-0.0752	-0.0795	-0.0838	-0.0757	-0.0789	-0.0967
24	-0.0283	-0.0398	-0.0557	-0.0543	-0.0629	-0.0773	-0.0341	-0.0384	-0.0672	-0.0341	-0.0672	-0.0730	-0.0644	-0.0687	-0.0470	-0.0513	-0.0672	-0.0687	-0.0644	-0.0686
25	-0.1034	-0.1207	-0.0928	-0.1121	-0.1207	-0.1293	-0.1034	-0.1078	-0.1101	-0.1034	-0.1250	-0.1250	-0.1164	-0.1207	-0.1164	-0.1207	-0.1250	-0.1207	-0.1164	-0.1379
26	0.2726	0.2553	0.2696	0.2612	0.2553	0.2426	0.2753	0.2628	0.2497	0.2671	0.2442	0.2510	0.2542	0.2526	0.2569	0.2540	0.2483	0.2512	0.2624	0.2381
27	-0.0983	-0.1156	-0.1026	-0.1121	-0.1156	-0.1293	-0.1034	-0.1026	-0.1199	-0.0983	-0.1250	-0.1250	-0.1113	-0.1207	-0.1113	-0.1207	-0.1250	-0.1207	-0.1164	-0.1379
28	-0.0427	-0.0620	-0.0449	-0.0513	-0.0536	-0.0643	-0.0385	-0.0386	-0.0621	-0.0447	-0.0642	-0.0621	-0.0556	-0.0578	-0.0535	-0.0599	-0.0642	-0.0599	-0.0535	-0.0792
29	-0.0849	-0.0990	-0.0892	-0.0904	-0.0990	-0.1107	-0.0849	-0.0892	-0.1095	-0.0880	-0.1033	-0.1064	-0.0916	-0.1052	-0.0947	-0.1052	-0.1095	-0.0990	-0.0978	-0.1194
30	-0.1034	-0.1176	-0.0923	-0.1090	-0.1176	-0.1231	-0.1004	-0.1016	-0.1219	-0.1034	-0.1219	-0.1219	-0.1133	-0.1114	-0.1133	-0.1176	-0.1219	-0.1145	-0.1133	-0.1348
31	-0.1034	-0.1207	-0.1078	-0.1121	-0.1207	-0.1293	-0.1034	-0.1078	-0.1250	-0.1034	-0.1250	-0.1250	-0.1164	-0.1207	-0.1164	-0.1207	-0.1250	-0.1207	-0.1164	-0.1379
32	0.0604	0.0456	0.0573	0.0517	0.0456	0.0382	0.0640	0.0585	0.0400	0.0653	0.0425	0.0400	0.0462	0.0419	0.0474	0.0419	0.0400	0.0443	0.0474	0.0283
33	-0.0458	-0.0630	-0.0549	-0.0592	-0.0630	-0.0716	-0.0602	-0.0453	-0.0721	-0.0506	-0.0769	-0.0673	-0.0635	-0.0678	-0.0683	-0.0534	-0.0673	-0.0678	-0.0635	-0.0947
34	-0.0198	-0.0332	-0.0260	-0.0322	-0.0408	-0.0362	-0.0179	-0.0203	-0.0413	-0.0217	-0.0452	-0.0433	-0.0308	-0.0389	-0.0327	-0.0389	-0.0413	-0.0313	-0.0346	-0.0505
35	-0.0897	-0.1070	-0.0941	-0.0847	-0.0796	-0.0882	-0.0761	-0.0804	-0.0976	-0.0761	-0.1113	-0.0976	-0.0890	-0.0796	-0.1027	-0.1070	-0.1113	-0.0933	-0.0753	-0.1242
36	-0.0271	-0.0404	-0.0314	-0.0344	-0.0470	-0.0530	-0.0245	-0.0288	-0.0461	-0.0258	-0.0461	-0.0474	-0.0374	-0.0404	-0.0335	-0.0431	-0.0474	-0.0431	-0.0361	-0.0577
37	-0.0757	-0.0929	-0.0800	-0.0843	-0.0929	-0.1015	-0.0757	-0.0800	-0.0972	-0.0757	-0.0972	-0.0972	-0.0886	-0.0929	-0.0886	-0.0929	-0.0972	-0.0929	-0.0886	-0.1102
38	-0.0209	-0.0381	-0.0240	-0.0307	-0.0381	-0.0456	-0.0221	-0.0229	-0.0425	-0.0221	-0.0483	-0.0413	-0.0350	-0.0405	-0.0303	-0.0381	-0.0425	-0.0370	-0.0315	-0.0542

Bibliography

[1] Alshammari, Riyad. "NIMS1 dataset", November 2011. URL http://web.cs.dal.ca/~riyad/Site/Download.html.

[2] Alshammari, Riyad and A. Nur Zincir-Heywood. "Investigating two different approaches for encrypted traffic classification". *Proceedings of the Sixth Annual IEEE Conference on Privacy, Security and Trust*, 156–166. 2008.

[3] Aurenhammer, Franz. "Voronoi diagrams—a survey of a fundamental geometric data structure". *ACM Computing Surveys*, 23(3):345–405, September 1991. ISSN 0360-0300.

[4] Brugger, S. Terry and Jedidiah Chow. *An assessment of the DARPA IDS Evaluation Dataset using Snort*. Technical report, University of California, Davis, November 2005.

[5] Callado, Arthur, Judith Kelner, Djamel Sadok, Carlos Alberto Kamienski, and Stênio Fernandes. "Better network traffic identification through the independent combination of techniques". *Journal of Network and Computer Applications*, 33(4):433–446, July 2010. ISSN 1084-8045.

[6] Cha, Sung-Hyuk. "Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions". *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4):300–307, November 2007.

[7] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey". *ACM Computing Surveys*, 41:15:1–15:58, July 2009. ISSN 0360-0300.

[8] Chen, Rung Ching, Kai-Fan Cheng, and Chia-Fen Hsieh. "Using Rough Set and Support Vector Machine for Network Intrusion Detection". *International Journal of Network Security & Its Applications*, 1(1):1–13, April 2009.

[9] Chmielewski, Andrzej and Sławomir T. Wierzchoń. "On the distance norms for detecting anomalies in multidimensional datasets". *Zeszyty Naukowe Politechniki Białostockiej*, 2:39–49, 2007.

[10] Cho, Jaeik, Kyuwon Choi, Taeshik Shon, and Jongsub Moon. "A Network Data Abstraction Method for Data Set Verification". *Secure and Trust Computing, Data Management and Applications*, volume 186 of *Communications in Computer and Information Science*, 54–62. Springer Berlin Heidelberg, June 2011. ISBN 978-3-642-22339-6.

[11] Cunningham, Pádraig. "A Taxonomy of Similarity Mechanisms for Case-Based Reasoning". *Knowledge and Data Engineering, IEEE Transactions on*, 21(11):1532–1543, November 2009. ISSN 1041-4347.

[12] Davis, Jonathan J. and Andrew J. Clark. "Data preprocessing for anomaly based network intrusion detection: A review". *Computers & Security*, 30(6–7):353–375, 2011. ISSN 0167-4048.

[13] Deb, Novarun, Manali Chakraborty, and Nabendu Chaki. "A State-of-the-Art Survey on IDS for Mobile Ad-Hoc Networks and Wireless Mesh Networks". Dhinaharan Nagamalai, Eric Renault, and Murugan Dhanuskodi (editors), *Advances in Parallel Distributed Computing*, volume 203 of *Communications in Computer and Information Science*, 169–179. Springer Berlin Heidelberg, September 2011. ISBN 978-3-642-24037-9.

[14] Deraman, Maznan, Abd Jalil Desa, and Zulaiha Ali Othman. "Public domain datasets for optimizing network intrusion and machine learning approaches". *Proceedings of the 2011 Conference on Data Mining and Optimization (DMO)*, 51–56. June 2011. ISSN 2155-6938.

[15] Deza, Michel Marie and Elena Deza. *Encyclopedia of distances*. Springer Verlag, 2009.

[16] Engen, Vegard, Jonathan Vincent, and Keith Phalp. "Exploring discrepancies in findings obtained with the KDD Cup '99 data set". *Intelligent Data Analysis*, 15:251–276, April 2011. ISSN 1088-467X.

[17] Freedman, David and Persi Diaconis. "On the histogram as a density estimator: L_2 theory". *Probability Theory and Related Fields*, 57:453–476, 1981. ISSN 0178-8051.

[18] García-Teodoro, Pedro, Jesús E. Díaz-Verdejo, Gabriel Maciá-Fernández, and Enrique Vázquez. "Anomaly-based network intrusion detection: Techniques, systems and challenges". *Computers & Security*, 28(1–2):18–28, 2009. ISSN 0167-4048.

[19] Grinstein, Georges, Sharon Laskowski, Graham Wills, and Bernice Rogowitz. "Information exploration shootout project and benchmark data sets (panel): evaluating how visualization does in analyzing real-world data analysis problems". *Proceedings of the Eighth IEEE Conference on Visualization*, 511–513. October 1997. ISBN 1-58113-011-2.

[20] Grinstein, Georges, Sharon Laskowski, Graham Wills, and Bernice Rogowitz. "The Information Exploration Shootout Data Set", November 2011. http://ivpr.cs.uml.edu/shootout/download.html.

[21] Gu, Guofei, Prahlad Fogla, David Dagon, Wenke Lee, and Boris Skorić. "Measuring intrusion detection capability: an information-theoretic approach". *Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security*, 90–101. March 2006. ISBN 1-59593-272-0.

[22] Gu, Guofei, Prahlad Fogla, David Dagon, Wenke Lee, and Boris Skoric. "Towards an Information-Theoretic Framework for Analyzing Intrusion Detection Systems".

Dieter Gollmann, Jan Meier, and Andrei Sabelfeld (editors), *European Symposium on Research in Computer Security*, volume 4189 of *Lecture Notes in Computer Science*, 527–546. Springer Berlin / Heidelberg, September 2006. ISBN 978-3-540-44601-9.

[23] Hancock, David L. and Gary B. Lamont. "Multi agent system for network attack classification using flow-based intrusion detection". *IEEE Congress on Evolutionary Computation*, 1535–1542. June 2011. ISSN Pending.

[24] Heidemann, John and Christos Papdopoulos. "Uses and Challenges for Network Datasets". *CATCH '09: Cybersecurity Applications Technology Conference For Homeland Security*, 73–82. March 2009.

[25] Iranmanesh, Seyed Mehdi, Mehdi Mohammadi, Ahmad Akbari, and Babak Nassersharif. "Improving Detection Rate in Intrusion Detection Systems Using FCM Clustering to Select Meaningful Landmarks in Incremental Landmark Isomap Algorithm". Qihai Zhou (editor), *Theoretical and Mathematical Foundations of Computer Science*, volume 164 of *Communications in Computer and Information Science*, 46–53. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-24999-0.

[26] Ji, Zhou and Dipankar Dasgupta. "Estimating the detector coverage in a negative selection algorithm". *Proceedings of the 2005 conference on Genetic and Evolutionary Computation*, GECCO '05, 281–288. ACM, New York, NY, USA, 2005. ISBN 1-59593-010-8.

[27] Ji, Zhou and Dipankar Dasgupta. "2-D Synthetic Datasets", December 2012. URL http://ais.cs.memphis.edu/index.php?page=datasets.

[28] Joseph, J.F.C. and A.A. Ghorbani. "VisVerND: Visual Verification of Network Traffic Dataset". *CNSR 2011: Proceedings of the Ninth Annual Communication Networks and Services Research Conference*, 56–62. May 2011.

[29] Jyothsna, V., V. V. Rama Prasad, and K. Munivara Prasad. "A Review of Anomaly based Intrusion Detection Systems". *International Journal of Computer Applications*, 28(7):26–35, August 2011.

[30] Kausar, Noreen, Brahim Belhaouari Samir, Azween Abdullah, Iftikhar Ahmad, and Mohammad Hussain. "A Review of Classification Approaches Using Support Vector Machine in Intrusion Detection". Azizah Abd Manaf, Shamsul Sahibuddin, Rabiah Ahmad, Salwani Mohd Daud, and Eyas El-Qawasmeh (editors), *Informatics Engineering and Information Science*, volume 253 of *Communications in Computer and Information Science*, 24–34. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-25462-8.

[31] Kenneally, E.E. and K. Claffy. "Dialing Privacy and Utility: A Proposed Data-Sharing Framework to Advance Internet Research". *IEEE Security & Privacy*, 8(4):31–39, July-August 2010. ISSN 1540-7993.

[32] Krishnamurthy, Balachander, Walter Willinger, Phillipa Gill, and Martin Arlitt. "A Socratic method for validation of measurement-based networking research". *Computer Communications*, 34(1):43–53, 2011. ISSN 0140-3664.

[33] Lakhina, Shilpa, Sini Joseph, and Bhupendra Verma. "Feature Reduction using Principal Component Analysis for Effective Anomaly-Based Intrusion Detection on NSL-KDD". *International Journal of Engineering Science and Technology*, 2(6):1790–1799, 2010.

[34] Lee, Suchul, Hyunchul Kim, Dhiman Barman, Sungryoul Lee, Chong-kwon Kim, Ted Kwon, and Yanghee Choi. "NeTraMark: a network traffic classification benchmark". *ACM SIGCOMM Computer Communication Review*, 41(1):22–30, January 2011. ISSN 0146-4833.

[35] Lippmann, Richard, Joshua W Haines, David J Fried, Jonathan Korba, and Kumar Das. "The 1999 DARPA off-line intrusion detection evaluation". *Computer Networks*, 34(4):579–595, 2000. ISSN 1389-1286.

[36] Lippmann, Richard, Joshua W Haines, David J Fried, Jonathan Korba, and Kumar Das. "DARPA Intrusion Detection Data Sets", November 2011. URL http://www.ll.mit.edu/mission/communications/ist/corpora/ideval/data/index.html.

[37] Mahoney, Matthew and Philip Chan. "An Analysis of the 1999 DARPA/Lincoln Laboratory Evaluation Data for Network Anomaly Detection". Giovanni Vigna, Christopher Kruegel, and Erland Jonsson (editors), *Recent Advances in Intrusion Detection*, volume 2820 of *Lecture Notes in Computer Science*, 220–237. Springer Berlin / Heidelberg, 2003. ISBN 978-3-540-40878-9.

[38] McHugh, John. "Testing Intrusion detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory". *ACM Transactions on Information and System Security*, 3(4):262–294, November 2000. ISSN 1094-9224.

[39] Moore, Andrew W. and Denis Zuev. "Internet traffic classification using bayesian analysis techniques". *SIGMETRICS Performance Evaluation Review*, 33:50–60, June 2005. ISSN 0163-5999.

[40] Moore, Andrew W. and Denis Zuev. "Dataset used in Internet Traffic Classification Using Bayesian Analysis Techniques", November 2011. URL https://www.cl.cam.ac.uk/research/srg/netos/nprobe/data/papers/sigmetrics/index.html. https://www.cl.cam.ac.uk/research/srg/netos/nprobe/data/papers/sigmetrics/index.html.

[41] Moradi, Farnaz, Magnus Almgren, Wolfgang John, Tomas Olovsson, and Philippas Tsigas. "On collection of large-scale multi-purpose datasets on internet backbone links". *Proceedings of the First Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, BADGERS '11, 62–69. ACM, New York, NY, USA, 2011. ISBN 978-1-4503-0768-0.

[42] Papadogiannakis, Antonis, Michalis Polychronakis, and Evangelos P. Markatos. "Improving the accuracy of network intrusion detection systems under load using selective packet discarding". *EUROSEC '10: Proceedings of the Third European ACM Workshop on System Security*, 15–21. 2010. ISBN 978-1-4503-0059-9.

[43] Pianka, E.R. "Niche overlap and diffuse competition". *Proceedings of the National Academy of Sciences*, 71(5):2141–2145, 1974.

[44] Pilli, E. S., R. C. Joshi, and R. Niyogi. "Data reduction by identification and correlation of TCP/IP attack attributes for network forensics". *ICWET '11: Proceedings of the International ACM Conference & Workshop on Emerging Trends in Technology*, 276–283. 2011. ISBN 978-1-4503-0449-8.

[45] Salama, Mostafa, Heba Eid, Rabie Ramadan, Ashraf Darwish, and Aboul Hassanien. "Hybrid Intelligent Intrusion Detection Scheme". AntÃşnio Gaspar-Cunha, Ricardo Takahashi, Gerald Schaefer, and Lino Costa (editors), *Soft Computing in Industrial Applications*, volume 96 of *Advances in Intelligent and Soft Computing*, 293–303. Springer Berlin / Heidelberg, 2011. ISBN 978-3-642-20504-0.

[46] Schuetz, Philipp and Amedeo Caflisch. "Efficient modularity optimization by multistep greedy algorithm and vertex mover refinement". *Physical Review E*, 77:046112, April 2008.

[47] Secure Ware. "ASHULA", April 2012. http://www.secure-ware.com/product/softashula.html.

[48] Sommer, Robin and Vern Paxson. "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection". *IEEE Symposium on Security and Privacy*, 305–316, 2010. ISSN 1081-6011.

[49] Song, Jungsuk, Hiroki Takakura, and Yasuo Okabe. "Kyoto University Benchmark Data dataset", November 2011. URL http://www.takakura.com/kyoto_data/. http://www.takakura.com/kyoto_data/.

[50] Song, Jungsuk, Hiroki Takakura, Yasuo Okabe, Masashi Eto, Daisuke Inoue, and Koji Nakao. "Statistical analysis of honeypot data and building of Kyoto 2006+ dataset for NIDS evaluation". *BADGERS '11: Proceedings of the First ACM Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, 29–36. 2011. ISBN 978-1-4503-0768-0.

[51] Soysal, Murat and Ece Guran Schmidt. "Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison". *Performance Evaluation*, 67(6):451–467, June 2010. ISSN 0166-5316.

[52] Staab, Steffen. "Ontologies and Similarity". Ashwin Ram and Nirmalie Wiratunga (editors), *Case-Based Reasoning Research and Development*, volume 6880 of *Lecture Notes in Computer Science*, 11–16. Springer Berlin / Heidelberg, 2011. ISBN 978-3-642-23290-9.

[53] Stolfo, Salvatore J. and Wenke Lee. "KDD Cup 1999 Data", November 2011. URL http://www.sigkdd.org/kddcup/index.php?section=1999&method=info.

[54] Tavallaee, M., E. Bagheri, Wei Lu, and A.A. Ghorbani. "A detailed analysis of the KDD CUP 99 data set". *CISDA 2009: IEEE Symposium on Computational Intelligence for Security and Defense Applications*, 1–6. July 2009.

[55] Tavallaee, M., E. Bagheri, Wei Lu, and A.A. Ghorbani. "The NSL-KDD Data Set", November 2011. URL http://www.iscx.ca/NSL-KDD/. http://www.iscx.ca/NSL-KDD/.

[56] Tavallaee, M., N. Stakhanova, and A.A. Ghorbani. "Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods". *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(5):516–524, September 2010. ISSN 1094-6977.

[57] Thomas, Ciza, Vishwas Sharma, and N. Balakrishnan. "Usefulness of DARPA Dataset for Intrusion Detection System Evaluation". *Proceedings of the SPIE 6973, 69730G (2008)*, Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security. The International Society for Optics and Photonics, March 2008.

[58] Varghese, George, J. Andrew Fingerhut, and Flavio Bonomi. "Detecting evasion attacks at high speeds without reassembly". *ACM SIGCOMM Computer Communication Review*, 36:327–338, August 2006. ISSN 0146-4833.

[59] Vigna, Giovanni. "Network intrusion detection: dead or alive?" *ACSAC '10: Proceedings of the 26th Annual Computer Security Applications Conference*, 117–126. ACM, New York, NY, USA, 2010. ISBN 978-1-4503-0133-6.

[60] Wang, Tuo, Shingo Mabu, Nannan Lu, and Kotaro Hirasawa. "A novel intrusion detection system based on the 2-dimensional space distribution of average matching degree". *Proceedings of the Annual Conference of the Society of Instrument and Control Engineers*, 2829–2834. September 2011. ISSN pending.

[61] Wolpert, D.H. and W.G. Macready. "No free lunch theorems for optimization". *Evolutionary Computation, IEEE Transactions on*, 1(1):67–82, April 1997. ISSN 1089-778X.

[62] Zhang, Min, Wolfgang John, K.C. Claffy, and Nevil Brownlee. "State of the art in traffic classification: a research overview". *PAM '09: Passive and Active Network Measurement Conference, Student Workshop*. 2009.

[63] Zhao, Kuo, Nurbol, Fei Ren, Jianfeng Chu, and Liang Hu. "Surveys on the Intrusion Tolerance System". Patrick Bond (editor), *Communications and Networking in China*, volume 26 of *Communications in Computer and Information Science*, 90–97. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-00205-2.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
21–03–2013	Master's Thesis	Oct 2011–Mar 2013

4. TITLE AND SUBTITLE

Network Intrusion Dataset Assessment

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Weller-Fahy, David J., Senior Master Sergeant, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-13-M-49

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Intentionally Left Blank.

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Research into classification using Anomaly Detection (AD) within the field of Network Intrusion Detection (NID), or Network Intrusion Anomaly Detection (NIAD), is common, but operational use of the classifiers discovered by research is not. One reason for the lack of operational use is most published testing of AD methods uses artificial datasets: making it difficult to determine how well published results apply to other datasets and the networks they represent. This research develops a method to predict the accuracy of an AD-based classifier when applied to a new dataset, based on the difference between an already classified dataset and the new dataset. The resulting method does not accurately predict classifier accuracy, but does allow some information to be gained regarding the possible range of accuracy. Further refinement of this method could allow rapid operational application of new techniques within the NIAD field, and quick selection of the classifier(s) that will be most accurate for the network.

15. SUBJECT TERMS

Computer networks, Intrusion detection, Dataset characterization, Network security, Network traffic representation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Lt Col Brett Borghetti, Asst. Professor (ENG)
U	U	U	UU	114	19b. TELEPHONE NUMBER *(include area code)* (937) 255-6565 ext. 4612