

School of Computer Science

Scientific Research & Literature

in Fulfilment of

SPEC9997

Maksymilian Drzezdzon

C15311966

Degree: TU060/2

Module Supervisor: Giancarlo Salton

Declaration of Ownership: I declare that the attached work is entirely my own and that all sources have been acknowledged: 🗹  
**Date: 2021/03/24**

Investigating differences in performance between SVM and DWD in Schizophrenia classification



***Maksymilian Drzezdzon***

*C15311966*

A dissertation submitted in partial fulfilment of requirements of Technological University Dublin for the degree of

M.Sc. in Computer Science

May 2022

# DECLARATION

I certify that this dissertation which I now submit for examination for the award of MSc in Computer Science, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the test of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Technological University Dublin and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute’s guidelines for ethics in research.

**Signed:** Maksymilian Drzezdzon

**Date:** 17/03/2022

# ABSTRACT

**Introduction**

There is an ever-growing emphasis on mental health with greater demands for already sparse mental health professionals. Applying machine learning as another criteria or test to satisfy a diagnosis for a mental illness such as schizophrenia would not only save valuable resources such as mental health practitioners time but also get patients the care they desperately need faster.

Young girls tend to be underdiagnosed with attention deficit hyperactivity disorder (ADHD) because how unalike its manifestation is when compared to boys, this levels out during adulthood, however having better tools for diagnosing disorders especially ones as serious as schizophrenia will greatly improve people’s quality of life and allow for medical intervention.

Schizophrenia is a disabling mental illness with huge time requirements to attain a diagnosis, a patient could avoid years of misdiagnosis/waiting if less severe or noticeable symptoms could be diagnosed/detected sooner through better diagnostic tools.

Obsessive compulsive disorder (OCD) and psychosis tend to be misdiagnosed as one another because the DSM-5 (diagnostic and statistical manual of mental disorders 5th edition) categorizes certain behavioral traits belonging to specific illnesses which in practice requires a lot more time and caution for evaluation, medications used to treat a patient experiencing psychosis which they may not have would actually exacerbate their OCD, leaving space for doubling down on/reinforcing a misdiagnosis that is then further mis/interpreted as psychosis after the fact.

There have been a few short comings in diagnosing of serious mental health disorders, there is no process to date that properly diagnoses dissociative identity disorder, despite it being acknowledged as a mental illness in the 1950s, research between then and now has been conducted but later found fraudulent or difficult to reproduce.

**Objective**

The goal of this study is to compare misclassification rates between state-of-the-art implementations of support vector machine (SVM) and distance weighed discrimination (DWD) applied to schizophrenia detection/classification through prototyping of classification models based on state-of-the-art research using preprocessed MRI and fMRI image modalities. Followed up by identifying areas for future work.

Give some indication of possible directions this work could go in the future.

Describe the sections of articles to come

**Results**

Body text - Body text Body text Body text Body text Body text Body text Body text

Keywords: Diagnosis Prediction, Schizophrenia Classification, Precision Psychiatry, Support Vector Machine, Distance Weight Discrimination

Word Count: **xxxxxx**

# ACKNOWLEDGEMENTS

Dataset used for this thesis is from a

Collection of this dataset was made at the [Mind Research Network](http://www.mrn.org/) under an NIH NIGMS Centers of Biomedical Research Excellence (COBRE) grant 5P20RR021938/P20GM103472 to Vince Calhoun (PI).

Notes for later:

Thank supervisor etc

REMOVE THIS LATER THIS WAS CUT FROM ABSTRACT – not really needed, add to intro instead

One study found and attained an accuracy of 70% with a 75% baseline when classifying schizophrenia on synthetic data with a DFNN model. In another study an accuracy of 84.4% was attained when using a SVM model on FNC between independent components extracted by ICA. Finally, one study achieved an accuracy of 87% on network maps extracted by ICA with an SVM model.

, provide scope of schizophrenia and MRI image formats and the digital imaging and communications (DICOM) standard which provides a central medium for imaging modalities.

Table of Contents

[DECLARATION i](#_Toc105258496)

[ABSTRACT ii](#_Toc105258497)

[ACKNOWLEDGEMENTS iii](#_Toc105258498)

[Chapter 1 – INTRODUCTION 1](#_Toc105258499)

[Background 1](#_Toc105258500)

[Research Problem 2](#_Toc105258501)

[Research Question 3](#_Toc105258502)

[Research Objectives 4](#_Toc105258503)

[Research Methodologies 5](#_Toc105258504)

[Scope and Limitations 5](#_Toc105258505)

[Document Outline 5](#_Toc105258506)

[Chapter 2 – LITERATURE REVIEW 7](#_Toc105258507)

[Introduction 7](#_Toc105258508)

[Approaches to solve the problem 8](#_Toc105258509)

[Gaps in Research 8](#_Toc105258510)

[Machine Learning in Mental Health 10](#_Toc105258511)

[HDLSS - TODO read paper 10](#_Toc105258512)

[Overview of Schizophrenia 10](#_Toc105258513)

[Functional Network Connectivity – FNC 10](#_Toc105258514)

[Sourced-Based Morphometry – SBM 10](#_Toc105258515)

[Evaluation 11](#_Toc105258516)

[Imaging Techniques 11](#_Toc105258517)

[DICOM 11](#_Toc105258518)

[MRI 11](#_Toc105258519)

[fMRI 11](#_Toc105258520)

[PET 11](#_Toc105258521)

[Chapter 3 – Design and Methodology 12](#_Toc105258522)

[Hypothesis 12](#_Toc105258523)

[Hypothesis testing 13](#_Toc105258524)

[Dataset 13](#_Toc105258525)

[Acknowledgement 13](#_Toc105258526)

[Dataset 13](#_Toc105258527)

[Data preparation 14](#_Toc105258528)

[Model training 14](#_Toc105258529)

[Feature selection 14](#_Toc105258530)

[RFE - Recursive Feature Elimination 14](#_Toc105258531)

[RRF - Regularized Random Forest 14](#_Toc105258532)

[LASSO - Last Absolute Shrinkage and Selection Operator 14](#_Toc105258533)

[Rpart - Decision Tree Rpart 14](#_Toc105258534)

[BORUTA 14](#_Toc105258535)

[3.4 15](#_Toc105258536)

[QVT 15](#_Toc105258537)

[Software/tools used 15](#_Toc105258538)

[Project Approach 15](#_Toc105258539)

[Research Assumptions 15](#_Toc105258540)

[Modeling 15](#_Toc105258541)

[Evaluation of results 15](#_Toc105258542)

[Chapter 4 – Results and Evaluation 16](#_Toc105258543)

[Research problem 16](#_Toc105258544)

[Implementation 16](#_Toc105258545)

[Results 17](#_Toc105258546)

[Explain results 18](#_Toc105258547)

[Conclusion 19](#_Toc105258548)

[Chapter 5 – Conclusion and Future Work 19](#_Toc105258549)

[Research Overview 19](#_Toc105258550)

[Research findings 19](#_Toc105258551)

[Future work and recommendations 20](#_Toc105258552)

[References 20](#_Toc105258553)

# Chapter 1 – INTRODUCTION

## Background

Machine learning is gaining popularity in many industries, one such industry is adopting healthcare which requires explainable results, some of the most powerful machine learning techniques available based on deep learning don’t produce explainable models.

A diagnosis for dissociative disorders and schizophrenia among other disorders rely on the interpretation of an assessment completed by a clinician, (HSE, 2021) this process is time consuming and prone to error especially when a patient does not exhibit severe symptoms such as delusions or the flattening of emotions, which is a condition where a person is unable to express emotions the same way other people might (Timothy J. Legg, 2017) Similar overlapping symptoms can cause a misdiagnosis and leave a patient astray for years before finally being diagnosed for instance with schizophrenia or a disassociation disorder.

Currently there is no non-invasive methods for diagnosing schizophrenia and no established biomarker for diagnosis besides using the process of elimination. **(Add citation form below, don’t forget)**

One way to reduce misdiagnosis is with the use of machine learning classification in conjunction with MRI and fMRI images. Once a biomarker for other illnesses such as depression or anxiety can be identified they can then be acknowledged and ‘omitted’ when searching for definitive biomarkers that help hone in on schizophrenia or disassociation disorders. Following the very traditional approach of elimination with a little more precision that could cater to different demographics as expressed in precision medicine, however that is outside the scope of this project.

An obstacle that occurs when one begins collecting data to analyze, apart from privacy concerns and difficulty obtaining such data due to regulations, schizophrenia only afflicts ~1% of the population making it very scarce. (mentalhelp.net, 2021) When working with images one can rotate them to synthesize more data for model training.

Machine learning algorithms and traditional statistical techniques being considered are support vector machine, linear discriminant analysis, multivariate analysis, neural networks, regression, k-nearest neighbor, k-means clustering and random forest. These approaches have been found to be most popular and effective when attempting to diagnose schizophrenia from MRI image data.

When talking about data, high dimensional low sample size (HDLSS) is typically not the first topic that comes to mind. However, in computer vision, complex mental illness classification such as schizophrenia via magnetic resonance imaging (MRIs), gene analysis, chemo-metrics handling it is one of many obstacles experts still wrestle with today (J. S. Marron, 2007) (Oh, 2020) (Chang Su, 2020).

(HDLSS) data is data that has more features than samples. When working with HDLSS data variations of classification methods such as support vector machine (SVM) and distance weighted discrimination (DWD) are used, (Delaram Sadeghi, 2021) each implementation tries to address shortcomings of its predeceasing method such as, neighbor-less nature (Cheema, 2015) or data piling in SVM (Jeongyoun Ahn, 2015) which is what prompted the creation of DWD. (J. S. Marron, 2007) What makes these challenges unique is that traditional classification approaches don't work and require tailored solutions. Most of these implementations don't have framework implementations yet which make experimentation difficult. (J. S. Marron, 2007) (Lock, 2020) Although deep learning models command impressive results (0.96 on an AUC) they tend to over fit and miss-classify younger cohorts with less severe symptoms (Oh, 2020) (Cortes, 2021) (Marjane Khodatars, 2021).

## Research Problem

This project investigates the use of modalities extracted from MRIs, methods used to acquire these coordinates were FNC & SBM, group independent component analysis and independent component analysis.

The most common way to acquire a diagnosis for schizophrenia is through a one-on-one assessment in conjunction with trial and error. There are no none invasive ways of diagnosing schizophrenia, especially when trying to identify patients with mild or lesser symptoms.

“If this can be accomplished through machine learning classification, what is the accuracy and error rate of the model, how does it compare to other methods?”

Support vector machine (SVM) is the state-of-the-art when classifying schizophrenia with HDLSS data. (Delaram Sadeghi, 2021) It’s a supervised learning algorithm (Boser, 1992) that utilizes a hyper-plane to separate two classes and assign them to either 'group'.

However, there is an issue if you don’t have enough data which would be the norm in a HDLSS setting, points mapped onto vectors end up or are close to 0 if not actually 0 making them identical, making it so that new samples can’t be properly grouped using a vanilla SVM classifier. Literature review – DWSVM seems to be the best – it tries to get the best of both worlds – cite, and research more

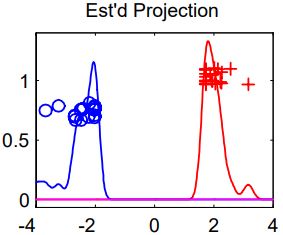


Figure 1: Example SVM data piling - Linear SVM, C = 1000, dimension = 39 (J. S. Marron, 2007)

This phenomenon is referred to as “*data-piling*”, (maybe move up before explaining what it is? It would read/flow more succinctly) which causes SVM to over-fit because it can’t use the data vector to differentiate points between the two classes resulting in tightly centered groups (J. S. Marron, 2007) (Jeongyoun Ahn, 2015) (AHN, 2010). This is where DWD and SVMs many implementations come into play. The solution to SVMs data piling is DWD.

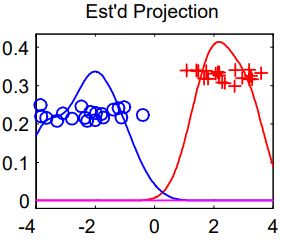


Figure 2: Example DWD no data piling - Distance Weighted Disc., dimension = 39 (J. S. Marron, 2007)

Looping back to the previous section regarding tailored solutions, there is a plethora of DWD implementations of which most don't have public code implementations appended or published making it difficult to compare approaches for experimentation.

## Research Question

"What are the differences in performance between implementations of SVM and DWD when classifying Schizophrenia using HDLSS data through fMRI/FNC features and sMRI/SBM loadings?"

## Research Objectives

This project is based on secondary research, using the Mind Research Network data-set supported by a systematic review of existing literature on SVM and DWD implementations for schizophrenia classification.

A deductive approach was used to form a hypothesis based on reviewed literature that will lead to an experiment from which metrics will be gathered to either confirm or refute the null hypothesis.

In this experiment the performance between SVM and DWD implementations will be examined using HDLSS data with an objective of using quantitative research methods via the development of classification models evaluated by (other error rate metrics, why use them?) F1 score, Log Loss, Categorical Cross entropy or AUC that will lead to the acceptance or rejection of the null hypothesis based on sample results gathered from model evaluation.

The first is to perform a deeper literature review and investigate how these coordinates were derived from fMRI and MRI images in order to get acquainted with the topic.

Finally build and evaluate a classification model prototype with appropriate metrics exploring the effect of feature selection on such a models performance using the data provided.

Aim:

To quantify the differences in misclassification rate through mean squared error between state-of-the-art implementations of SVM and DWD for HDLSS data in mental illness detection/classification

Objectives:

* Implement SVM, DWD, BayesianDWD, FunctionalDWD, kernDWD, SparseDWD, DWDLargeR, penalizedSVM, SVM-Maj, SparseSVM, WeightSVM and ParallelSVM
* Setup feature selection methods for each algorithm implementation
* Gather metrics from each model on unfiltered data vs subsets gathered from feature selection
* Repeat feature selection and training if time allows with QVT implementation (Wanwan Zheng, 2022)
* Format results into table with regards to which algorithm was used, feature selection improvement criteria & QVT impact if implemented

## Research Methodologies

A review of previous literature was carried out, scoping the area of machine learning applications to mental health diagnostics and/or classification. A dataset has been found and cited for this undertaking. Each algorithm will have several feature selection methods used in conjunction with QVT which seems to be a significant augmentation for HDLSS data. (Wanwan Zheng, 2022) As a means of evaluation 20% of the data-set used will be used to measure performance.

Model performance will be critiqued highlighting possible avenues for future PhD work.

## Scope and Limitations

The scope of this project is to build classification models comparing their performance after tuning given the available timeframe. This project is limited to the dataset available inheriting any plausible assumptions made during gathering data from patients.

* Limited to reading in preprocessed MRI modalities
* Preprocessing MRIs from schizconnect is time consuming and out of scope for the allocated time frame
* Severity of schizophrenia could be a factor in model sensitivity, this severity wasn’t recorded

## Document Outline

Chapter 2 - Literature review

Reviewing existing literature in mental illness classification using (update this with whats in chp 2). Surface level overview of schizophrenia and motivations of this research along with imaging techniques. Review methods used in acquiring data, FNC and SBM.

Chapter 3 - Design and methodology

This chapter focusses on how the project is conducted, the dataset used, its features, data preparation and the proposed solution. Model evaluation is discussed along with a description of how the project will be conducted providing methods employed to test the hypothesis.

The hypothesis is stated followed by how it will be tested, later a description of the dataset used is provided, how the data was initially acquired by the original research team, how it was prepared and explored for this project. A brief list of software, languages and tools used.

Chapter 4 - Results and evaluation

This chapter focusses on summarizing results of the experiment and evaluate the proposed method. The chapter concludes with a discussion on strengths and limitations of the proposed solution highlighting potential improvements and areas for future research.

Chapter 5 - Conclusion

A summary of key findings, conclusions and areas for future research.

# Chapter 2 – LITERATURE REVIEW

## Introduction

This chapter covers all required components of this project. A quick overview of schizophrenia will be provided along with a review of papers and methods used to synthesize the dataset from MRI and fMRI images. This will be augmented with a short list of popular classification methods used in disorder classification in respect to current literature. Finishing with an account on machine learning application in mental disorder classification.

The state-of-the-art research focuses on two categories of methods, machine learning predominately via SVM (Olivier Chapelle, 2002) and deep learning techniques (Delaram Sadeghi, 2021). The downfall of deep learning (DL) especially in the HDLSS space is due to it being data hungry and the challenge here is on the opposite side of the spectrum. (Oh, 2020) Most research leans towards traditional machine learning because the performance doesn't drop off as sharply as it does with DL when more diverse cohorts which is very difficult to capture in a HDLSS data-set but also with such a rare condition. Gathering enough labeled data spread evenly more or less on age groups would in itself be a milestone.

The main differences in synthesizing a deep learning model compared to traditional machine learning is the feature extraction, as illustrated in figure 4 (Delaram Sadeghi, 2021) (A. Jović, 2015) (Mwangi, 2014).

Most used feature extraction techniques can be grouped into categories which are (Delaram Sadeghi, 2021) (Federico Calesella, 2021) statistical that encompass the mean, variance and standard deviation. Textural features, a more specialized method in medical imaging utilized in deep learning (Georgiadis, 2008), non-linear feature extraction uses neuro-imaging modalities with a publicly available framework hosted on GitHub called NeAT (Adrià Casamitjana, 2020) (Pinotsis, 2014) and graph models (Yizhen Xiang, 2020) (Sugai Liang, 2018) are used in a HDLSS setting to combat how computationally expensive feature selection is with larger HDLSS data-sets. (Zhihong Zhang, 2011)

Feature reduction narrows down to principal component analysis as a common approach when leveraging traditional machine learning methods. (Delaram Sadeghi, 2021) (Svante Wold, 1987) The large majority of these studies focus around SVM or deep learning methods utilizing methods such as convolutional neural networks, recurrent neural networks, generative adversarial networks and the list goes on. Other more sophisticated implementations such as CADS are used, however those require huge compute power along with profound expertise in the topic to tune these models (Delaram Sadeghi, 2021).

These can be validated using methods such as T-tests and analysis of variance (ANOVA) for feature selection in neuro-medical images. In all studies reviewed accuracy, sensitivity and recall were used to compare performance, however (Delaram Sadeghi, 2021) (Mwangi, 2014)

Current implementations of DWD and SVM have been narrowed down to functional DWD (Sang, 2021), kernDWD (Wang, 2018), sparse DWD \cite{Wang2016}, Weighed SVM \cite{XuleiYang}, sparse SVM \cite{Zhou2021}, MajSVM \cite{Yip2019TheMA}, penalizedSVM \cite{Becker2009, SINGH2021100014}, DWDlargeR \cite{Lam2016FastAF}.

### Approaches to solve the problem

Current approaches to tackle HDLSS data are different implementations that build upon SVM, DWD or deep learning methods. Current state of the art methods seem to use deep learning (Delaram Sadeghi, 2021) (Qingbo Yin, 2020) (Huang, 2012) (Artemiou, 2021). As mentioned previously DWD has seen little adoption over SVM or deep learning with research beginning to pile up over the last few years, over a decade after its initial proposal in 2004/2007. (J. S. Marron, 2007)

### Gaps in Research

J. S. Marron authored the idea of DWD (J. S. Marron, 2007) to tackle SVMs data piling problem. DWD is sensitive to the sample size ratio between classes denoted by the intercept term β, this is a problem because when taking into consideration the differences between cohorts age, stage of schizophrenia, type of schizophrenia among other intricacies that make it hard to diagnose and distinguish between, once and if accounted for, this can cause/causes a class imbalance, this limitation is blunted with refined implementations of DWD such as wDWD and DWSVM but at a significantly higher computational cost (Qingbo Yin, 2020).

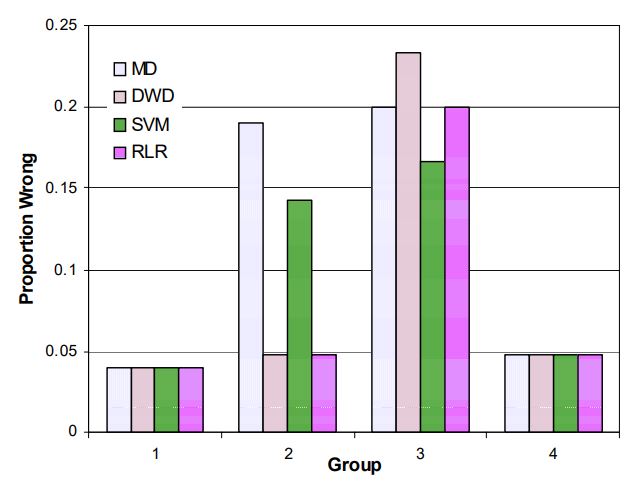


Figure 3: Error rates for gene expression data (J. S. Marron, 2007)

Despite how well DWD and its various implementations cater to HDLSS data (Huang, 2012) (Artemiou, 2021) it’s not widely used in most state of the art research (Delaram Sadeghi, 2021) even with its competitive performance (J. S. Marron, 2007) , this seems like an application/implementation gap where DWD and its following implementations are only being picked up on more recently. The idea would be to compare accuracy of different SVM and DWD implementations with available implementations online and compile a set of results in the 20 weeks allocated for this project.

In order to better understand the methodological gap an investigation between both methods and their subsequent implementations that follow should be undergone.

An alternative study could be conducted in the future to implement other variations of SVM and DWD in python or R libraries as most don't have ready to use implementations.

## Machine Learning in Mental Health

Common applications of ML in healthcare encompass detection and diagnosis, prognosis, treatment and support, public health applications, research and clinical administration.

These can be further segregated into pre-diagnosis screening tools and risk models that identify individual’s predisposition or susceptibility to develop mental health conditions. (Adrian B. R. Shatte, 2019)

In this project the disorder being examined is schizophrenia with mention of disassociation as they have overlapping symptomology. A goal here could be to better differentiate the two.

Other avenues worth exploring are the use of NLP to detect onset of schizophrenia, similar was accomplished with Alzheimer’s disease. (Elif Eyigoz, 2020) Potentially analyzing social media activity could also be a viable path for future work depending on data availability.

### HDLSS - TODO read paper

High dimensionality low sample size data (Dan Shen, 2016)

## Overview of Schizophrenia

Schizophrenia is not only a serious mental illness for the afflicted but costly to the healthcare system, using up already scare healthcare resources. Finding better ways to help with diagnostics would not only help the individual but also alleviate stress put on clinicians.

Schizophrenia was first identified by Emil Kraepelin in 1896 under the name dementia praecox. (R M Ion, 2002) It’s very difficult to diagnose schizophrenia due to the fact that it overlaps with many other illnesses or conditions such as disassociation and psychosis.

## Functional Network Connectivity – FNC

FNC are correlation values that summarize the overall connection between independent brain maps over time through correlation in statistical analysis. FNC describe patterns of brain function. (Elena A. Allen, 2014) It’s important to note that this data refers to the state at a given point in time, meaning patients must be in the same state when this kind of data is being gathered. This is done through MRI, fMRI, EEG or MEG. In this project these values were acquired from fMRI from schizophrenic patients and healthy controls at rest (rs-MRI) with group independent component analysis. (Elena A Allen 1, 2011)

## Sourced-Based Morphometry – SBM

SBM loadings are weights of brain maps gathered from gray matter concentration maps using independence component analysis. These values are also derived from MRI scans and represent a patient’s brain structure. (Judith M Segall 1, 2012) The goal behind such metrics is that they provide cognitive capability for each region of the brain through statistical analysis.

## Evaluation

Sadasdasdasd

## Imaging Techniques

### DICOM

Digital imaging and communications in medicine standard is a data interchange protocol for biomedical image format’s structure. (W. Dean Bidgood, 1997)

### MRI

Magnetic resonance imaging is used in radiology to take none invasive images of brain and brain stem structure. (Michael Harkin, 20217)

### fMRI

Functional magnetic resonance images are similar to MRIs but depict the changes in blood oxygen levels. It’s been used in conjunction with statistical methods for classification for concluding inferences about brain states. (Glover, 2012)

### PET

A positron emission tomography scan uses a chemical/dye containing tracers which can be viewed by a PET scanner. In brain disorder classification this can be used to detect levels of glucose similar to SBM weights, PET scans can be used to inspect regions of the brain that use more or less glucose. Currently its used for Alzheimer’s disease and depression. (Brian Krans, 2018) depression is sometimes diagnosed in tandem with schizophrenia.

Maybe write here about the use of AI/ML with Imaging techniques ?? otherwise this is redundant.

# Chapter 3 – Design and Methodology

This chapter will focus on the experiment used to evaluate models and accept or reject the null hypothesis. Several sets of models are implemented for both SVM and DWD. All models will be tested on the Mind Research Network MLSP 2014 dataset. Data preparation, feature selection is described along with the train/test process. Further more a second pass of high preforming models will be selected and retrained with the QVT (citation) algorithm used to enhance feature selection for HDLSS models.

The F1 score, Cohens Kappa, McNamar’s Test P-value, PPV (positive prediction value), NPV (negative prediction value), log loss and AUC will be collected for each model’s assessment along with the public and private Kaggle scores.

Finally, a state-of-the-art implementation of DW-SVM would be implemented and compared to the rest of the models.

This study aims to investigate the effectiveness of DWD through classification accuracy metrics such as F1 score, Log Loss, Categorical Cross entropy and AUC to gauge its effectiveness when applied to HDLSS data for schizophrenia classification compared to SVM on the same data-set.

This project focuses on secondary research with an empirical approach, using quantitative methods and deductive reasoning that will grant or refute the null hypothesis.

## Hypothesis

**Null Hypothesis:**

If SVM implementations are used to classify schizophrenic patient’s vs healthy controls using fMRI/FNC features (correlation values that summarize connection between brain maps over time) and sMRI/SBM loadings (weights of brain maps derived from gray matter concentration of all subjects) then the mean difference between the F1 score, Log Loss and AUC of the top 5 SVM models and top 5 DWD models will not be statistically significant.

**Alternate Hypothesis:**

If SVM implementations are used to classify schizophrenic patient’s vs healthy controls using fMRI/FNC features (correlation values that summarize connection between brain maps over time) and sMRI/SBM loadings (weights of brain maps derived from gray matter concentration of all subjects) then the mean difference between the F1 score, Log Loss and AUC of the top 5 SVM models and top 5 DWD models will be statistically significant.

## Hypothesis testing

The type of statistical test is dependent on the sampling technique used, for each model trained, in this scenario, using k-fold validation will mean that estimated metrics are dependent, it further limits the possibility for truly independent samples (Brownlee, Statistical significance tests for comparing machine learning algorithms, 2019) (Brownlee, A gentle introduction to estimation statistics for Machine Learning, 2019).

Using a Non-parametric tests such as a paired t-test Wilcoxon singed rank test can be used to validate the hypothesis, albeit it holds less statistical power. Alternatively, estimation statistics can be used such as effect size, interval estimation, confidence intervals or meta-analysis that can be used to accept or reject the null hypothesis.

There are a few ways to do this, such as the use of PCA to lower the number of features by their highest eigen-values for dimension reduction or factor analysis followed by a MANOVA for a global hypothesis test.

This is preferred if data is of Gaussian distribution validated or otherwise in objective 8, this rule has been slightly relaxed over time \cite{muller\_stewart\_2006}. Another option is PCA or factor analysis, for PCA to be successful it requires a “simple co-variance structure, at least asymptotically” (Yueh-Yun Chi, 2013).

Factor analysis is sensitive to an unequal ratio of observations to variables which also holds true for PCA. There are other specialized alternatives, however more time needs to be allocated to identify better suited method/s.

Reject the null hypothesis (Ho) if p < 0.05 from the above hypothesis test using a MANOVA with the use of PCA/factor analysis is found.

## Dataset

### Acknowledgement

Collection of this dataset was made at the [Mind Research Network](http://www.mrn.org/) under an NIH NIGMS Centers of Biomedical Research Excellence (COBRE) grant 5P20RR021938/P20GM103472 to Vince Calhoun (PI)

### Dataset

Describe dataset here

### Data preparation

The FNC and SBM train sets are loaded in and merged by the ID column alongside the labels dataset that contains the binary value for whether a patient has schizophrenia by ID. Each row is a patient with 410 features spread between SBM and FNC. The experiment was run against several subsets of the original dataset that were filtered by select feature selection algorithms along with removing linear dependencies on remaining features before training each model. Each subset was saved to a csv for ease of use as experiments were carried out. The goal of this approach was to remove noise and reduce

### Model training

Each model was first trained using a train control parameter using repeated cross validation 10 times with 10 folds, allowing parallel processing if a model supports it to save time before saving the final predictions. This was done to get a sense of initial parameters that would be selected for a model. Afterword a grid would be constructed using values that encompass the initial values selected to try find better settings. Typically, this approach increased the models AUC by 0.03-0.14 and NPV/PPV by 0.10. Models such as sparse DWD need more manual tunning as lambda values would go below 0.

Make a diagram?

Talk about different kernels used?

## Feature selection

A brief description of feature selection algorithms used.

### RFE - Recursive Feature Elimination

### RRF - Regularized Random Forest

### LASSO - Last Absolute Shrinkage and Selection Operator

### Rpart - Decision Tree Rpart

### BORUTA

### 3.4

### QVT

Algorithms augmented tried with QVT

## Software/tools used

The R programming language in conjunction with R studio was used, every library used is available via the c-ran package manager.

## Project Approach

The scope is limited to examining classification techniques such as SVM, DWD and their implementations over a period of ~13 weeks applied to FNC/SBM correlation values gathered from HDLSS data from the Mind Research Network’s Schizophrenia Dataset consisting of 35,432 observations gathered from 162 patients and 169 healthy controls. The Aim of this study is to derive differences between implementations and their classification accuracy via the F1 score

## Research Assumptions

sadasdasdasas

## Modeling

sadasdasdas

## Evaluation of results

sadasdasdas

# Chapter 4 – Results and Evaluation

## Research problem

sadasdasdas

## Implementation

sadsdasdasdas

## Results

**Table 1: SVM**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithm used | Model Metrics | | | | | | Kaggle Evaluation Scores | | Feature Selection |
| SVM | F1 | Cohens Kappa | Mcnemar’s  Test P-value | PPV | NPV | AUC | Private | Public | Method used |
| SVM - Poly | 0.6364 | 0.3548 | 0.72367 | 0.7 | 0.6667 | 0.68 | 0.79 | 0.630 | N/A |
|  | 0.6957 | 0.4373 | 1 | 0.7273 | 0.7413 | 0.72 | 0.679 | 0.616 | RFE |
|  | 0.72 | 0.441 | 1 | 0.6923 | 0.75 | 0.72 | 0.779 | 0.777 | RRF |
|  | 0.555 | 0.3464 | 0.077 | 0.83 | 0.631 | 0.67 | 0.63 | 0.58 | LASSO |
|  | 0.5833 | 0.1987 | 1 | 0.5833 | 0.615 | 0.6 | 0.679 | 0.54 | Rpart |
|  | 0.8696 | 0.7588 | 0.7588 | 0.91 | 0.857 | 0.88 | 0.741 | 0.71 | BORUTA |
| SVM - Radial | 0.5263 | 0.2671 | 0.1824 | 0.7143 | 0.6111 | 0.63 | 0.708 | 0.643 | N/A |
|  | 0.6957 | 0.4373 | 1 | 0.7273 | 0.7143 | 0.72 | 0.737 | 0.713 | RFE |
|  | 0.74 | 0.44 | 0.44969 | 0.667 | 0.8 | 0.72 | 0.751 | 0.741 | RRF |
|  | 0.6 | 0.3506 | 0.2888 | 0.75 | 0.647 | 0.71 | 0.741 | 0.571 | LASSO |
|  | 0.6364 | 0.3548 | 0.72367 | 0.7 | 0.6667 | 0.68 | 0.669 | 0.58 | Rpart |
|  | 0.833 | 0.6795 | 1 | 0.833 | 0.8462 | 0.84 | 0.751 | 0.705 | BORUTA |
| SVM - Linear | 0.5714 | 0.2718 | 0.505 | 0.667 | 0.625 | 0.63 | 0.684 | 0.54 | N/A |
|  | 0.7273 | 0.5161 | 0.68309 | 0.8 | 0.733 | 0.76 | 0.674 | 0.616 | RFE |
|  | 0.6667 | 0.5098 | 0.041 | 1 | 0.733 | 0.75 | 0.779 | 0.714 | RRF |
|  | 0.88 | 0.7604 | 1 | 0.8462 | 0.9167 | 0.88 | 0.54 | 0.59 | LASSO |
|  | 0.5 | 0.0385 | 1 | 0.5 | 0.539 | 0.52 | 0.679 | 0.607 | Rpart |
|  | 0.8462 | 0.682 | 0.617 | 0.7857 | 0.91 | 0.84 | 0.718 | 0.763 | BORUTA |
| SVM - Sig | 0.5714 | 0.2718 | 0.505 | 0.667 | 0.625 | 0.63 | 0.79 | 0.696 | N/A |
|  | 0.6957 | 0.4373 | 1 | 0.7273 | 0.7143 | 0.72 | 0.774 | 0.652 | RFE |
|  | 0.72 | 0.4409 | 1 | 0.6923 | 0.75 | 0.72 | 0.746 | 0.808 | RRF |
|  | 0.75 | 0.5192 | 1 | 0.75 | 0.7692 | 0.76 | 0.713 | 0.643 | LASSO |
|  | 0.72 | 0.4409 | 1 | 0.6923 | 0.75 | 0.72 | 0.723 | 0.634 | Rpart |
|  | 0.818 | 0.678 | 0.617 | 0.9 | 0.8 | 0.84 | 0.674 | 0.741 | BORUTA |
|  | ***Note****: RRF – Regularized Random Forest, RFE – Recursive Feature Elimination, LASSO – Last Absolute Shrinkage and Selection Operator, Rpart – Decision Tree Rpart, PPV – Positive Predicted Value, NPV – Negative Predicted Value* | | | | | | | | |

**Table 2: DWD**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DWD Kernel | Model Metrics | | | | | | Kaggle Evaluation Scores | | Feature Selection |
| DWD | F1 | Cohens Kappa | Mcnemar’s  Test P-value | PPV | NPV | AUC | Private | Public | Method used |
| Linear | 0.7 | 0.513 | 0.22 | 0.875 | 0.7059 | 0.75 | 0.644 | 0.584 | N/A |
|  | 0.8462 | 0.6815 | 0.617 | 0.785 | 0.909 | 0.84 | 0.794 | 0.61 | RFE |
|  | 0.8182 | 0.677 | 0.617 | 0.9 | 0.8 | 0.84 | 0.702 | 0.714 | RRF |
|  | 0.761 | 0.595 | 0.371 | 0.888 | 0.75 | 0.8 | 0.651 | 0.629 | LASSO |
|  | 0.63 | 0.35 | 0.723 | 0.7 | 0.667 | 0.68 | 0.756 | 0.562 | Rpart |
|  | 0.4 | 0.257 | 0.0076 | 1 | 0.59 | 0.62 | 0.653 | 0.53 | BORUTA |
| Radial | 0.7143 | 0.367 | 0.2888 | 0.625 | 0.777 | 0.69 | 0.641 | 0.642 | N/A |
|  | 0.7692 | 0.522 | 0.683 | 0.7143 | 0.8182 | 0.76 | 0.817 | 0.741 | RFE |
|  | 0.75 | 0.5192 | 1 | 0.75 | 0.769 | 0.76 | 0.717 | 0.607 | RRF |
|  | 0.7273 | 0.51 | 0.683 | 0.8 | 0.733 | 0.76 | 0.741 | 0.571 | LASSO |
|  | 0.6 | 0.3506 | 0.288 | 0.75 | 0.647 | 0.67 | 0.75 | 0.53 | Rpart |
|  | 0.833 | 0.6795 | 1 | 0.833 | 0.846 | 0.84 | 0.642 | 0.642 | BORUTA |
| Poly | 0.6667 | 0.359 | 1 | 0.6667 | 0.6923 | 0.68 | 0.684 | 0.633 | N/A |
|  | 0.88 | 0.76 | 1 | 0.8462 | 0.9167 | 0.88 | 0.679 | 0.669 | RFE |
|  | 0.9565 | 0.9196 | 1 | 1 | 0.928 | 0.96 | 0.641 | 0.642 | RRF |
|  | 0.667 | 0.5098 | 0.04 | 1 | 0.6842 | 0.75 | 0.6358 | 0.571 | LASSO |
|  | 0.6957 | 0.434 | 1 | 0.727 | 0.714 | 0.72 | 0.717 | 0.531 | Rpart |
|  | 0.8182 | 0.6774 | 0.617 | 0.9 | 0.8 | 0.84 | 0.669 | 0.718 | BORUTA |
| Sparse | 0.69 | 0.36 | 0.72 | 0.642 | 0.727 | 0.68 | 0.607 | 0.58 | N/A |
|  | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | RFE |
|  | 0.72 | 0.4409 | 1 | 0.6923 | 0.75 | 0.72 | 0.679 | 0.736 | RRF |
|  | 0.7 | 0.513 | 0.22 | 0.875 | 0.7 | 0.75 | 0.70 | 0.48 | LASSO |
|  | 0.4 | 0.257 | 0.0076 | 1 | 0.59 | 0.62 | 0.717 | 0.642 | Rpart |
|  | 0.47 | 0.26 | 0.0455 | 0.8 | 0.6 | 0.63 | 0.664 | 0.59 | BORUTA |
|  | ***Note****: RRF – Regularized Random Forest, RFE – Recursive Feature Elimination, LASSO – Last Absolute Shrinkage and Selection Operator, Rpart – Decision Tree Rpart, PPV – Positive Predicted Value, NPV – Negative Predicted Value* | | | | | | | | |

Poly – rfe everything

Poly – rprt no lin normal model

Poly – rrf norm has better AUC but grid has better Kaggle score

Boruta, lasso no lin for lin dwd

## Explain results

asdasdasdasdas

## Conclusion

asdasdasdasdasdasdas

# Chapter 5 – Conclusion and Future Work

## Research Overview

Sadasdas

## Research findings

asdasdasdas

## Future work and recommendations

Asdasdasdasdasdas

# References

A. Jović, K. B. (2015). A review of feature selection methods with applications. *Electronics and Microelectronics (MIPRO)*, pp. 1200-1205, doi: 10.1109/MIPRO.2015.7160458.

Adrià Casamitjana, V. V. (2020). NeAT: a Nonlinear Analysis Toolbox for Neuroimaging. *neuroinformatics* , 517–530 (2020). https://doi.org/10.1007/s12021-020-09456-w.

Adrian B R Shatte, D. M. (2019 July). Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine*, https://doi.org/10.1017/S0033291719000151.

Adrian B. R. Shatte, D. M. (2019). Machine learning in mental health: A systematic scoping review of methods and applications. *Psychological medicine*, 49(9), 1426–1448. https://doi.org/10.1017/S0033291719000151.

AHN, J. &. (2010). The maximal data piling direction for discrimination. *Biometrika*, 97(1), 254–259. http://www.jstor.org/stable/27798914.

Artemiou, A. R. (2021). Sufficient dimension reduction based on distance-weighted discrimination. *Scandinavian Journal of Statistics*, https://doi.org/10.1111/sjos.12484.

Boser, G. V. (1992). *Support Vector Machines.* Boston, MA: Springer.

Brian Krans, C. S. (2018, September 17). *What Is a PET Scan?* Retrieved from healthline: https://www.healthline.com/health/pet-scan#purpose

Brownlee, J. (2019, August 8). *A gentle introduction to estimation statistics for Machine Learning*. Retrieved from machinelearningmastery.com: https://machinelearningmastery.com/estimation-statistics-for-machine-learning/?fbclid=IwAR0mdkFN\_hFzvAtIinRey2LueMdUR8oAeQ3hX3pqz\_UHHJBuc9-iwKEn\_ug

Brownlee, J. (2019, August 8). *Statistical significance tests for comparing machine learning algorithms*. Retrieved from machinelearningmastery.com: https://machinelearningmastery.com/statistical-significance-tests-for-comparing-machine-learning-algorithms

Chang Su, Z. X. (2020). Deep learning in mental health outcome research: a scoping review. *Translational Psychiatry*, 10(1), 116. https://doi.org/10.1038/s41398-020-0780-3.

Cheema, M. S. (2015). High dimensional low sample size activity recognition using geometric classifiers. *Digit. Signal Process*, 61–69. https://doi.org/10.1016/j.dsp.2015.03.019.

CONNOLLY, M. (2020, Jan 17). *adhdireland*. Retrieved from Easy-to-Miss ADHD Symptoms in Girls: https://adhdireland.ie/what-adhd-looks-like-in-girls/

Cortes, J. &.-R. (2021). Going deep into schizophrenia with artificial intelligence. *Schizophrenia Research*, 10.1016/j.schres.2021.05.018.

Dan Shen, H. S. (2016). The Statistics and Mathematics of High Dimension Low Sample Size Asymptotics. *Statistica Sinica*, 26(4), 1747–1770. 10.5705/ss.202015.0088.

Delaram Sadeghi, A. S.-D. (2021). An Overview on Artificial Intelligence Techniques for Diagnosis of Schizophrenia Based on Magnetic Resonance Imaging Modalities: Methods, Challenges, and Future Works. *Advanced Researches In Biomedical Engineering Lab.*, https://doi.org/10.48550/arXiv.2103.03081.

Elena A Allen 1, E. B. (2011). A baseline for the multivariate comparison of resting-state networks. *Frontiers in systems neuroscience*, 5, 2. https://doi.org/10.3389/fnsys.2011.00002.

Elena A. Allen, E. D. (2014). Tracking Whole-Brain Connectivity Dynamics in the Resting State . *Cerebral Cortex*, 663–676, https://doi.org/10.1093/cercor/bhs352.

Elif Eyigoz, S. M. (2020). Linguistic markers predict onset of Alzheimer's disease. *The Lancet*, https://doi.org/10.1016/j.eclinm.2020.100583.

Federico Calesella, A. T. (2021). A comparison of feature extraction methods for prediction of neuropsychological scores from functional connectivity data of stroke patients. *Brain Informatics* , https://doi.org/10.1186/s40708-021-00129-1.

Georgiadis, P. &. (2008). Improving brain tumor characterization on MRI by probabilistic neural networks on non-linear transformation of textural features. *Computer methods and programs in biomedicine*, 89. 24-32. 10.1016/j.cmpb.2007.10.007 .

Glover, G. H. (2012). Overview of Functional Magnetic Resonance Imaging. *Neurosurgery clinics of North America*, 10.1016/j.nec.2010.11.001.

HSE. (2021, April 4). *Diagnosis Schizophrenia*. Retrieved from HSE.ie: https://www2.hse.ie/conditions/mental-health/schizophrenia/schizophrenia-diagnosis.html#:~:text=There's%20no%20single%20test%20for,the%20result%20of%20other%20causes.

Huang, H. &. (2012). R/DWD: Distance Weighted Discrimination for Classification, Visualization and Batch Adjustment. *Bioinformatics* , 28. 1182-3. 10.1093/bioinformatics/bts096.

J. S. Marron, M. J. (2007). Distance-Weighted Discrimination. *Journal of the American Statistical Association*, 102(480), 1267–1271. http://www.jstor.org/stable/27639976.

Jeongyoun Ahn, Y. J. (2015). Sparse HDLSS discrimination with constrained data piling. *Computational Statistics & Data Analysis*, 74-83, https://doi.org/10.1016/j.csda.2015.04.006.

Judith M Segall 1, E. A. (2012). Correspondence between structure and function in the human brain at rest. *Frontiers in neuroinformatics*, 6, 10. https://doi.org/10.3389/fninf.2012.00010.

Lock, E. F. (2020). Bayesian Distance Weighted Discrimination. *arXiv*, 10.48550/ARXIV.2010.03111.

Marjane Khodatars, A. S. (2021). Deep learning for neuroimaging-based diagnosis and rehabilitation of Autism Spectrum Disorder: A review. *Computers in Biology and Medicine*, 104949. https://doi.org/10.1016/j.compbiomed.2021.104949.

mentalhelp.net. (2021, April 4). *Schizophrenia Symptoms, Patterns and Statistics and Patterns*. Retrieved from mentalhelp.net: https://www.mentalhelp.net/schizophrenia/statistics/#:~:text=Worldwide%20about%201%20percent%20of,this%20year%20around%20the%20world.

Michael Harkin, S. H. (20217, March 29). *What is a head MRI?* Retrieved from healthline: https://www.healthline.com/health/head-mri

Mwangi, B. T. (2014). A Review of Feature Reduction Techniques in Neuroimaging. *Neuroinformatics*, 229–244 (2014). https://doi.org/10.1007/s12021-013-9204-3.

Oh, J. a.-L.-U.-H. (2020). Identifying Schizophrenia Using Structural MRI With a Deep Learning Algorithm. *Frontiers in Psychiatry*, 10.3389/fpsyt.2020.00016.

Olivier Chapelle, V. V. (2002). Choosing Multiple Parameters for Support Vector Machines. *Springer Machine Learning*, 131–159. https://doi.org/10.1023/A:1012450327387.

Paul A. Tiffin, L. W. (2018). Rise of the machines? Machine learning approaches and mental health: opportunities and challenges. *The British Journal of Psychology* , 213, 509–510. http://doi.org/10.1192/bjp.2018.105.

Pinotsis, D. &. (2014). Extracting novel information from neuroimaging data using neural fields. *Nonlinear Biomedical Physics*, 2. 5. 10.1186/1471-2202-15-S1-O4. .

Qingbo Yin, E. A. (2020). Population-guided large margin classifier for high-dimension low-sample-size problems. *Elsevier*, https://doi.org/10.1016/j.patcog.2019.107030.

R M Ion, M. D. (2002). The British reaction to dementia praecox 1893-1913. Part 1. *Hist Psychiatry*, 10.1177/0957154X0201305103.

Reinders, A. M. (2019). Aiding the diagnosis of dissociative identity disorder: pattern recognition study of brain biomarkers. *The British journal of psychiatry* , 215(3), 536–544. https://doi.org/10.1192/bjp.2018.255.

Sang, P. &. (2021). A reproducing kernel Hilbert space framework for functional data classification. *researchgate*, researchgate.net/publication/349914131\_A\_reproducing\_kernel\_Hilbert\_space\_framework\_for\_functional\_data\_classification.

Sugai Liang, R. V. (2018). Neurocognitive Graphs of First-Episode Schizophrenia and Major Depression Based on Cognitive Features. *Neuroscience bulletin*, 10.1007/s12264-017-0190-6.

Svante Wold, K. E. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 37-52, doi.org/10.1016/0169-7439(87)80084-9.

Timothy J. Legg, P. (2017, August 4). *What is Flat Affect?* Retrieved from healthline: https://www.healthline.com/health/flat-affect#:~:text=Flat%20affect%20is%20a%20condition,affect%20shows%20no%20facial%20expressions.

W. Dean Bidgood, J. M. (1997). Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging. *Journal of the American Medical Informatics Association*, 10.1136/jamia.1997.0040199.

Wang, B. Z. (2018). Another look at distance-weighted discrimination. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1369-7412. https://doi.org/10.1111/rssb.12244.

Wanwan Zheng, M. J. (2022). Improving the Performance of Feature Selection Methods with Low-Sample-Size Data . *The Computer Journal*, https://doi.org/10.1093/comjnl/bxac033.

Yizhen Xiang, J. W.-X. (2020). Schizophrenia Identification Using Multi-View Graph Measures of Functional Brain Networks. *Frontiers in Bioengineering and Biotechnology*, 10.3389/fbioe.2019.00479 .

Yueh-Yun Chi, K. E. (2013). Two-Step Hypothesis Testing When the Number of Variables Exceeds the Sample Size. *Communications in Statistics - Simulation and Computation*, 1113-1125. 10.1080/03610918.2012.659819.

Yuhui Du1, Z. F. (06 August 2018). Classification and Prediction of Brain Disorders Using Functional Connectivity: Promising but Challenging. *frontiers in Neuroscience*.

Zhihong Zhang, E. R. (2011). *A Graph-Based Approach to Feature Selection.* Berlin: Springer.