

School of Computer Science

Scientific Research & Literature

in Fulfilment of

SPEC9997

Maksymilian Drzezdzon

C15311966

Degree: TU060/1

Module Coordinator: John Gilligan

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.

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.

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

**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

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

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.

# LIST OF FIGURES

# ACRONYMS

|  |  |
| --- | --- |
| ML | Machine Learning |
| SVM | Support Vector Machine |
| KNN | K-nearest neighbour |
| NN | Neural Network |
| rs-MRI | Rest State Magnetic Resonance Imaging |
| rs-fMRI | Rest State Functional Magnetic Resonance Imaging |
| D.I.D | Dissassociative Identity Disorder |
|  |  |

Do I want this here?

Table of Contents

[DECLARATION i](#_Toc100428737)

[ABSTRACT ii](#_Toc100428738)

[ACKNOWLEDGEMENTS iii](#_Toc100428739)

[LIST OF FIGURES iv](#_Toc100428740)

[ACRONYMS v](#_Toc100428741)

[Chapter 1 – INTRODUCTION 1](#_Toc100428742)

[Background 1](#_Toc100428743)

[Research Problem 2](#_Toc100428744)

[Research Objectives 2](#_Toc100428745)

[Research Methodologies 2](#_Toc100428746)

[Scope and Limitations 2](#_Toc100428747)

[Document Outline 3](#_Toc100428748)

[Chapter 2 – LITERATURE REVIEW 4](#_Toc100428749)

[Introduction 4](#_Toc100428750)

[Overview of Schizophrenia 4](#_Toc100428751)

[Functional Network Connectivity – FNC 4](#_Toc100428752)

[Sourced-Based Morphometry – SBM 4](#_Toc100428753)

[Machine Learning in Mental Health 5](#_Toc100428754)

[SVM 5](#_Toc100428755)

[DWD 5](#_Toc100428756)

[HDLSS 5](#_Toc100428757)

[Evaluation 5](#_Toc100428758)

[Imaging Techniques 5](#_Toc100428759)

[DICOM 5](#_Toc100428760)

[MRI 5](#_Toc100428761)

[fMRI 6](#_Toc100428762)

[PET 6](#_Toc100428763)

[Chapter 3 – Design and Methodology 6](#_Toc100428764)

[Hypothesis 6](#_Toc100428765)

[Hypothesis testing 6](#_Toc100428766)

[Dataset 6](#_Toc100428767)

[Feature selection 6](#_Toc100428768)

[How it was gathered 6](#_Toc100428769)

[Data preparation 6](#_Toc100428770)

[Data exploration 6](#_Toc100428771)

[Methods used on data 7](#_Toc100428772)

[Software/tools used 7](#_Toc100428773)

[Project Approach 7](#_Toc100428774)

[Research Assumptions 7](#_Toc100428775)

[Modeling 7](#_Toc100428776)

[Evaluation of results 7](#_Toc100428777)

[Chapter 4 – Results and Evaluation 7](#_Toc100428778)

[Research problem 7](#_Toc100428779)

[Implementation 8](#_Toc100428780)

[Results 8](#_Toc100428781)

[Explain results 8](#_Toc100428782)

[Conclusion 8](#_Toc100428783)

[Chapter 5 – Conclusion and Future Work 9](#_Toc100428784)

[Research Overview 9](#_Toc100428785)

[Research findings 9](#_Toc100428786)

[Future work and recommendations 9](#_Toc100428787)

[References 10](#_Toc100428788)

# 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 can be difficult when a patient does not exhibit more 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.

## Research Problem

Usually, MRI images are used for detecting mental health disorders. This project investigates the use of coordinates provided in the dataset extracted from MRI images, the methods used to acquire these coordinates were FNC & SBM, group independent component analysis and independent component analysis.

Other than one on one assessments in conjunction with trial and error there are no none invasive ways of diagnosing schizophrenia, especially when trying to identify patients with milder symptoms. If this can be accomplished, what is the accuracy of the new model, how does it compare to other methods, how accurate is it if applied to other mental health disorders?

## Research Objectives

There are a few objectives for this project. 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. Then research how one could better select features from such a dataset. Dimension reduction can be used for feature selection to potential identify biomarkers and record any positive effects when making predictions. 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.

## Research Methodologies

A review of previous literature will be 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. Furthermore, the institution that provided this dataset will be contacted for any possible research that was completed thereafter.

This project will focus on the work that came before it in the form of scoping reviews and state-of-the-art research papers, qualifying this and secondary research. Algorithm and statistical techniques that will be used have already been listed above, empirical research will be used to evaluate an array of metrics such as accuracy, F1-score and Cohens Kappa.

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

## Scope and Limitations

The scope of this project is to build a classification model and identify its performance after ideal configurations are identified. This project is limited to the dataset available inheriting any plausible issues or wrong assumptions made from a lack of clinical expertise when interpreting results or the data itself.

## Document Outline

Chapter 2 - Literature review

Existing literature covering mental illness classification using machine learning is reviewed and discussed. The aim is to become acquainted with current research and apply machine learning classification algorithms to processed fMRI and MRI images. Along with what led to this research.

Chapter 3 - Design and methodology

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

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.

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.

## 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. Describe the rest here

## 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.

## 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.

### SVM

sadasdasd

### DWD

asdasdasd

### HDLSS

asdasdasdasd

## 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.

# Chapter 3 – Design and Methodology

## Hypothesis

Adasdasdasd

## Hypothesis testing

adasdasdas

## Dataset

sdasdadasdas

## Feature selection

asdasdasdasd

## How it was gathered

asdasdasdasd

## Data preparation

Asdasdasdasd

## Data exploration

Asdasdasdas

## Methods used on data

asdasdasdasd

## Software/tools used

asdasdasdas

## Project Approach

asdasdasdas

## Research Assumptions

sadasdasdasas

## Modeling

sadasdasdas

## Evaluation of results

sadasdasdas

# Chapter 4 – Results and Evaluation

## Research problem

sadasdasdas

## Implementation

sadsdasdasdas

## Results

asdasdasdasd

## Explain results

asdasdasdasdas

## Conclusion

asdasdasdasdasdasdas

# Chapter 5 – Conclusion and Future Work

## Research Overview

Sadasdas

## Research findings

asdasdasdas

## Future work and recommendations

Asdasdasdasdasdas

# References

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.

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

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.

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

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.

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.

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.

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

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.

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.

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.

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