Aus der Klinik für Psychiatrie und Psychotherapie
der Ludwig-Maximilians-Universität München
Direktor: Prof. Dr. med. Peter Falkai

Application of support vector machines in psychiatry: the case of multimodal data integration and stratification based on sex and gender

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von

Maria Fernanda Urquijo Castro

aus

Barranquilla, Kolumbien

Jahr

2025

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. med. Nikolaos Koutsouleris
Mitberichterstatter:	Prof. Dr. Thomas Geyer
	Prof. Dr. Michael Ingrisch
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	20.01.2025

Affidavit

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Eidesstattliche Versicherung

Urquijo Castro, Maria Fernanda

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

Application of support vector machines in psychiatry: the case of multimodal data integration and stratification based on sex and gender

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 30.01.2025 — Maria Fernanda Urquijo Castro

Ort, Datum — Unterschrift Doktorandin bzw. Doktorand

Confirmation of Congruency

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Promotionsbüro
Medizinische Fakultät

MMRS

Confirmation of congruency between printed and electronic version of the doctoral thesis

Urquijo Castro, Maria Fernanda

Surname, first name

I hereby declare, that the submitted thesis entitled:

Application of support vector machines in psychiatry: the case of multimodal data integration and stratification based on sex and gender

is congruent with the printed version both in content and format.

Munich, 30.01.2025
place, date

Maria Fernanda Urquijo Castro
Signature doctoral candidate

Table of contents

List of abbreviations

BAC	Balanced accuracy
Cx	Cisgender
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
GMV	grey matter volume
HC	Healthy Controls
ICD-11	International Classification of Diseases, 11th Revision
ML	Machine learning
MDD	Major Depressive Disorder
rsfMRI	resting state functional magnetic resonance imaging
SABV	Sex as a biological variable
SAGER	Sex and Gender Equity in Research
sMRI	structural magnetic resonance imaging
SVM	Support Vector Machine
SZ	Schizophrenia
Tx	Transgender

List of publications

Paper I

Cabral, C., Kambeitz-Ilankovic, L., Kambeitz, J., Calhoun, V. D., Dwyer, D. B., Von Saldern, S., **Urquijo MF.,** Falkai P. & Koutsouleris, N. (2016). **Classifying schizophrenia using multimodal multivariate pattern recognition analysis: evaluating the impact of individual clinical profiles on the neurodiagnostic performance**. *Schizophrenia bulletin*, *42*(suppl_1), S110-S117. https://doi.org/10.1093/schbul/sbw053

Number of citations by the time of thesis' submission according to Google Scholar:

ISI Web of Knowledge: Schizophrenia Bulletin
Impact factor 2016: 7.575

Paper II

Baldinger-Melich, P.*, **Urquijo Castro, M. F.*,** Seiger, R., Ruef, A., Dwyer, D. B., Kranz, G. S., Klöbl, M., Kambeitz, J., Kaufmann, U., Windischberger, C., Kasper, S., Falkai, P., Lanzenberger, R. & Koutsouleris, N. (2020). **Sex matters: a multivariate pattern analysis of sex-and gender-related neuroanatomical differences in cis-and transgender individuals using structural magnetic resonance imaging.** *Cerebral Cortex*, 30(3), 1345-1356. https://doi.org/10.1093/cercor/bhz170 *Authors contributed equally
Number of citations by the time of thesis' submission according to Google Scholar:

ISI Web of Knowledge: Cerebral Cortex
Impact factor 2020: 5.375

Contribution to publications

Paper I

In paper I, we investigated the benefits of integrating multiple neuroimaging data modalities when classifying individuals with schizophrenia versus healthy controls using support vector machines. The open-source dataset COBRE (http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html) was used for this purpose. Here, I contributed to the definition of the main research question after taking part in lively discussions, and performed the analyses by applying machine learning techniques using the NeuroMiner toolbox (http://proniapredictors.eu/neurominer/index.html). Specifically, I implemented Support Vector Machines to create models on two different data modalities derived with magnetic resonance imaging, i.e., structural and resting-state functional brain models, from which the decision scores for each modality were extracted. I proceeded to use these scores to train a third superordinate model, from which the multimodal class membership was obtained. Thereafter, I contributed to the reporting of the methods section, the redaction of the manuscript and the revision process leading to the final publication.

Paper II

Paper II, is the end-product of my extensive efforts and collaboration with Associate Professor Dr. med. univ. et scient. med. Pia Baldinger-Melich from the Medical University of Vienna. After defining a detailed plan, I took the lead in employing Support Vector Machines to analyse neuroimaging data previously obtained by Dr. Baldinger-Melich's workgroup on a sample of individuals with gender dysphoria. Specifically, I developed reliable support vector machine models utilizing grey matter volume to identify the best distinguishing patterns between male and female, both cis- and transgender, individuals. This included preparing the data, defining meaningful parameters to enter the algorithm, and running the analyses. Additionally, I employed a crossover approach by applying the cisgender model to the transgender group and vice versa. Furthermore, I conducted additional analyses by applying the developed models to an external dataset of individuals with a Major Depressive Disorder diagnosis. I led the manuscript drafting process including the conception and development of the introduction, methods, results, figures, discussion, and conclusions and finally, I was primarily responsible for integrating feedback from reviewers, addressing their comments, and rewriting the final version of the manuscript which was accepted for publication.

Introduction

Precision psychiatry is an emerging field that proposes a paradigm shift in the way psychiatric disorders are diagnosed and treated. Unlike traditional approaches that rely on a one-fits-all model, precision psychiatry emphasizes the need to consider individual differences in biology, environment, and lifestyle to create diagnostic and treatment procedures that are most effective for each individual.

Although classical statistical methods have been instrumental in advancing the understanding of mental disorders, some pitfalls have been identified. These methods often rely on group-level analyses, use models that do not account for the relationship between variables and make an a priori selection of variables that perhaps are not the ones better explaining the studied phenomena (Dwyer et al., 2018). Machine learning (ML), a part of artificial intelligence that mimics principles of human pattern learning (Spicer & Sanborn, 2019), can address these limitations. Specifically Support Vector Machines (SVMs) can handle high-dimensional data and capture complex, non-linear relationships between variables, providing a more nuanced understanding of psychiatric disorders. They excel in integrating diverse data sources, a critical requirement for precision psychiatry.

Particularly in neuroimaging, information from diverse neurobiological data such as functional and structural MRI provide important information about psychiatric illness, such as schizophrenia and other psychoses (Porter et al., 2023; Schultz et al., 2012). However, integrating these diverse data modalities in a meaningful manner remains a significant challenge.

Furthermore, precision psychiatry aims at finding the best-suited diagnostic and prognostic tools for each individual. Yet, biological sex as well as gender identity have been mostly treated as nuisance variables when creating mental health models. (Cahill, 2014, 2017; Kurth et al., 2021). Initiatives like the ‘Sex as a biological variable’ (SABV) policy (Clayton & Collins, 2014) and the Sex and Gender Equity in Research (SAGER) guidelines (De Castro et

al., 2016), have prompted the field to consider, collect, characterize, and communicate sex-based data to improve applicability of scientific results.

Considering these aspects, the main objective of this thesis is to evaluate, within the framework of precision psychiatry and neuroimaging, the advantage of applying ML, specifically SVMs when 1) integrating various data modalities versus the use of single neuroimaging data types in the search of psychoses biomarkers and 2) aiming for a sex/gender-based stratification in precision psychiatry. By focusing this thesis on the application of SVMs, we aim to highlight their pivotal role in advancing the field of precision psychiatry and improving healthcare strategies for individuals with mental disorders.

General concepts of machine learning

Machine learning can be defined as “a computational strategy that automatically determines (i.e., learns) methods and parameters to reach an optimal solution to a problem rather than being programmed by a human a priori to deliver a fixed solution” (Dwyer et al., 2018). While the invention of such data-driven approaches dates to the 1930s, it is only in the past few decades that they have been integrated in the medical sciences. So far, mostly cancer research has succeeded to translate results into clinically usable tools (Esteva et al., 2017; Yu et al., 2016). In psychiatry, ML holds promise to improve differential diagnostic procedures, treatment response, relapse, and early recognition of mental illness (Rutledge et al., 2019) (Dwyer et al., 2018).

Depending on the input provided to the algorithm, ML models can be categorized as supervised or unsupervised. Supervised ML algorithms, such as Support Vector Machines, work with a priori labelled datasets, for example male/female or patient/control) and identify the characteristics that best distinguishes between these labels. In contrast, unsupervised models, e.g. clustering, aim at finding the *natural* grouping of unlabelled information based on patterns of similarities of the data points (Jain, 2010). This thesis focuses on Support Vector Machines (SVM) as these are one of the most frequently utilized supervised techniques and were employed in the two research projects in this work.

For ML-based models to be reliably in clinical practice, they must provide clinically meaningful predictions that are applicable to individuals that were not part in the model-building phase. However, most models published so far in psychiatry do not generalize well to new unseen data, i.e., they are overfitted to the original sample (Rosen et al., 2021). To guarantee an adequate level of generalizability, the current gold-standard strategy is cross-validation, consisting of a random splitting of the data into smaller groups, then using some data subsets for model training and the remaining parts for performance testing (Dwyer et al., 2018). The training data is used to optimize parameters such as feature selection strategies (meaning which feature combination is most informative and reliable) and the number of allowed misclassified cases (Dwyer et al., 2018). After several iterations of training and testing the models in different data subsets, model performance can be evaluated in terms of sensitivity, specificity, and accuracy. Sensitivity identifies true positives (individuals having disorder X correctly labelled by the algorithm as having disorder X), specificity identifies true negatives (individuals not having disorder X correctly labelled as not having disorder X), and accuracy represents the number of correct classifications with respect to all cases. Alternatively, a

refined version of accuracy, i.e. Balanced Accuracy (BAC), can be used. BAC deals with unbalanced samples and is obtained by adding the sensitivity and specificity and then dividing the result by two (Dwyer et al., 2018). Which model is finally selected as “winner” is determined, for instance, on the highest average BAC achieved (Koutsouleris et al., 2020; Schaffer, 1993).

Robust ML approaches can help psychiatry overcome some obstacles faced by clinicians and researchers, as it promotes the search of individualized answers beyond group-wise differences. It allows for the comprehensive analysis of large and complex datasets with many variables simultaneously, which can reveal previously unknown inter-variable relationships. And finally, ML promotes the translation of scientific results based on naturalistic data into real life settings (Mechelli & Vieira, 2020).

Support Vector Machines (SVMs)

SVMs are supervised machine learning techniques that learn to discriminate between predefined classes or groups based on different features (Cortes & Vapnik, 1995) such as variables from a questionnaire, voxels from MRI or single questions from a clinical questionnaire. For this learning process to occur, "support vectors" need to be identified. Support vectors are data points lying closest to an optimal separating boundary, known as the optimal separating hyperplane (Orru et al., 2012) (see figure 1). The optimal separating hyperplane is embedded into a margin which can allow for a degree of misclassification. In accordance with the Vapnik-Chervonenkis statistical learning theory (Vapnik, 1999), SVMs aim for robust performance by achieving high accuracy not only during the training phase but also in testing scenarios. The main objective is to develop models that are reproducible and generalizable to new data. Thus, robust SVMs models are created after testing different parameter combinations that are ideally optimized by the algorithm without human interference. Moreover, SVM models require enough features per case (i.e. variables per study subject) while also ensuring that the necessary variance/variability is available across cases (Pisner & Schnyer, 2020)

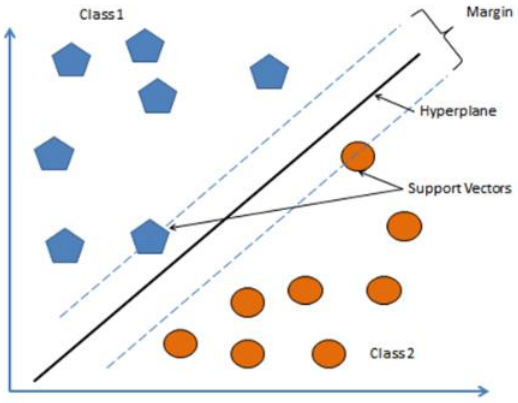

Figure 1 Diagram of a Support Vector Machine

SVMs have gained significant traction in neuroimaging research within psychiatry as they excel at handling high-dimensional datasets, making them particularly suited to handle the complex and multifactorial nature of psychiatric disorders and the brain.

Compared to classical statistics, machine learning approaches provide results in terms of single subject characteristics (Dwyer et al., 2018). Moreover, it is suited to integrate diverse data modalities such as functional and structural MRI to obtain a fuller picture or the studied psychiatric illness (Chen et al., 2021; Porter et al., 2023; Schultz et al., 2012), and using machine learning can help to provide proof of the relevance of factors such as sex and gender in psychiatric models. This has been ignored in spite of the large evidence describing major sex-based differences in psychiatric

disorders such as schizophrenia (Egloff et al., 2018; Li et al., 2016; Riecher-Rossler et al., 2018); depression and anxiety (Schuch et al., 2014) (Eid et al., 2019) and gender dysphoria (Feder et al., 2017; Gonzalez et al., 2017).

Multimodal data integration

Given that psychiatric disorders are highly complex and multifactorial, attempts to understand them require collecting, analysing and reporting data from different sources from each single person. Therefore, since it is usual during a clinical MRI scanning session to acquire information on more than one modality such as structural and resting state functional MRI, it is a logical step to take advantage of their complementarity beyond the radiological boundaries (Plitman et al., 2020). Merging neuroimaging modalities has been consistently shown to increase accuracy when characterizing diverse disorders such as autism (Mueller et al., 2013), anorexia (Cha et al., 2016) and Alzheimer's disease (Weiler et al., 2015).

Specifically, in the context of research on schizophrenia and other psychoses, multimodal data integration is particularly valuable, given that the illness has been associated to both structural and functional brain abnormalities (Calhoun & Sui, 2016; Porter et al., 2023; Schultz et al., 2012). Moreover, even though schizophrenia is a severe chronic mental health condition that affects more than 20 million individuals worldwide and has been associated with significant impairment of quality of life, no biomarker has been established yet. To date, diagnosis is established on the basis of the presence or central characteristics such as hallucinations, delusions, cognitive disorganization among others, (American Psychiatric Association, 2013), but the neurobiological underpinnings remain unclear.

Structural MRI studies have identified alterations in brain volume and cortical thickness (Fusar-Poli & Meyer-Lindenberg, 2016), while resting-state functional MRI has revealed disruptions in brain connectivity patterns from early to chronical stages of the disease (Allebone et al., 2018; Argyelan et al., 2015; O'Neill et al., 2019). These neurobiological changes suggest that psychosis involves complex interactions between brain structure and function. Therefore, integrating data from both structural and functional MRI is crucial to comprehensively grasp the neural underpinnings of psychosis. Although this approach poses significant statistical challenges, it provides a more holistic view of an individual's neurobiological state, thereby enhancing the understanding of psychopathological mechanisms and improving diagnostic and prognostic accuracy

Our 2016 study marked a significant milestone in this journey. By demonstrating the benefits of integrating multimodal neuroimaging data, we paved the way for newer, more sophisticated approaches. Based on our findings, the field has moved towards finding a brain-based predictome of mental illness, (Rashid & Calhoun, 2020). Moreover, similar approaches integrating multiple modalities have been applied to bipolar disorder (Li et al., 2020) and finally other classifiers (Ridge, Lasso, Random Forests, and Gradient boosting) have been tested in the schizophrenia research field (Salvador et al., 2019). Thus, as we continue to refine the integration of diverse data modalities and apply cutting-edge machine learning techniques in neuroimaging, we move closer to the goal of personalized mental health care.

Sex-Based differences in psychiatry

Besides multimodal data integration, another strategy to better tailor individualized psychiatric care programs is sex-based stratification. In biological research, *sex* is recognized as the biological variable referring to biological systems which serve to assign individuals to either the male or female category at birth based on the appearance of the external genitalia (American Psychiatric Association (APA), 2015). It is inherent to humans and regarded as binary (female/male) although clinical conditions related to sex development exist in around 0.018% of the population (García-Acero et al., 2020; Sax, 2002).

Studies assessing neuroanatomical sex-based differences consistently hint at larger brains in male vs. female individuals (Ritchie et al., 2018; Ruigrok et al., 2014) but greater gyrification in female compared to male individuals in frontal and parietal regions (Luders et al., 2004), leading to an overall comparable cortical surface in both sexes. More fine-grained structures such as the hippocampus, amygdala, the hypothalamus, and the basal ganglia also appear to be volumetrically different between sexes (Lotze et al., 2019). However, most of these findings are surrounded by strong inconsistencies with respect to magnitude, exact location, and some studies even report opposite direction of effects (i.e., volume increase vs. decrease in certain areas for either sex). Some studies have claimed that the brain cannot be categorized as sexually dimorphic since the volume of most areas is largely overlapping between the two sexes, with only a few small parts found to be exclusively "male" or "female." (Joel et al., 2015). Nevertheless, by employing SVMs to search for whole brain patterns, accuracies larger than 80% have been reported when studying sexual dimorphism in healthy individuals (Chekroud et al., 2016) (Rosenblatt, 2016). These results indicate that sex is reflected in rather subtle and widespread across the brain instead of limited to particular features and that, given some methodological limitation of classical univariate statistics, ML methods might be a better strategy to characterize these differences.

Sex stratification in gender dysphoria and major depressive disorder

In contrast to sex, gender or gender identity refers to the sociocultural norms and expectations associated to having either male or female sex phenotypical traits, which entails a subjective position of alignment or mismatch with the own biological sex. Neither sex nor gender identity are to be confused with the concept of sexual orientation, which defines a persons' sexual attraction to individuals of one or more genders such as hetero-, homo-, bi- or asexuality.

When an individual's gender identity aligns with their biological sex, they are considered cisgender (Cx). Conversely, when there is a divergence, individuals may identify as transgender (Tx) or gender diverse (Bouman & Arcelus, 2017; Thorne et al., 2020). Cases where this misalignment leads to significant distress are classified under gender dysphoria in the DSM-5 and gender incongruence in the ICD-11. Although it is not clear which proportion of the world's population is transgender or gender diverse the estimation ranges between 0.1 and 2% (Zhang et al., 2020), showing an upward trend (Nolan et al., 2019). A larger ratio of individuals with male biological sex experiencing gender dysphoria has been consistently reported across studies (Zucker, 2017).

Neuroanatomical research on transgender individuals has highlighted alterations in brain structures typically associated with male or female sex compared to cisgender individuals of the same biological sex (Dekaban & Sadowsky, 1978; Luders et al., 2009; Savic et al., 2010). Specifically, regions commonly recognized as sexually dimorphic appear partially feminized in male-to-female transgender individuals and partially masculinized in female-to-male transgender individuals (Guillamon et al., 2016). However, these findings predominantly stem from animal models or post-mortem studies, which often focus on predefined brain regions and employ diverse statistical methods. This variability complicates comparisons and limits individual-level insights. Additionally, many studies fail to consider confounding factors such as hormone therapy status and sexual orientation.

The second study in this dissertation was among the first to use SVMs to evaluate the effects of gender identity on brain structure while controlling for these confounding variables. Our findings have paved the way for further exploration of the neuroanatomical substrates of gender identity (Flint et al., 2020; Kurth et al., 2022) and has led researchers to advocate for innovative neuroradiological techniques for studying specifically the transgender brain (Stowell et al., 2022).

In addition, we applied our cisgender and transgender biological sex classification models to a sample of patients with depression. This allowed us to investigate how biological sex and gender identity interact with depression. Despite the DSM-5 no longer classifying gender dysphoria as a disorder (Davy & Toze, 2018), it has been associated with an increased vulnerability towards the development of psychopathology (de Freitas et al., 2020; Garg et al., 2018) in particular, mood disorders. This increased vulnerability can be attributed to a combination of psychosocial and biological factors. The distress and impairment associated with gender dysphoria, often exacerbated by societal stigma, discrimination, and lack of social support, significantly contribute to the elevated incidence of depression, reaching up to 62% in transgender women (Rotondi et al., 2012; Witcomb et al., 2018) .

Neurobiological evidence suggests shared pathways between gender dysphoria and depression, including dysregulation in the hypothalamic-pituitary-adrenal (HPA) axis and alterations in neurochemical systems involving serotonin and dopamine (Healy, 2015; Müller & Holsboer, 2005). Brain regions such as the insula and anterior cingulate cortex, which play crucial roles in self-perception and emotional regulation, exhibit structural and functional differences in individuals with gender dysphoria compared to those without (Feusner et al., 2017; Reed et al., 2023). These areas are also implicated in the pathophysiology of depression. Moreover, sex hormones like estrogen and testosterone influence mood regulation and neuroplasticity (Fernandez-Guasti et al., 2012), potentially exacerbating depressive states in individuals with gender dysphoria, especially under the unique social stressors faced by minority groups.

The application of machine learning in studying sex differences and gender dysphoria can uncover complex interactions between brain structure, gender identity, and mental health, leading to more accurate diagnostics and personalized treatments. For the Tx population, such advancements promise improved mental health outcomes and better quality of life. Continued progress in this field is essential, as it not only enhances scientific knowledge but also promotes social equity and well-being for gender-diverse individuals. By leveraging advanced scientific methods, we can address the unique challenges faced by the transgender community and foster a more inclusive society.

Rationale and aim

This thesis aims to investigate the potential benefits of using ML approaches to integrate multiple data modalities as well as stratify results based on sex/gender in psychiatric research. We hypothesized that:

1) Classifying individuals with schizophrenia from healthy controls machine learning on brain volumetric data is feasible and integrating different data modalities increases accuracy compared to single modalities (paper I).

2) The signature of biological sex neuroanatomical level is clear (model accuracy above 75%) but it relates to a pattern instead of single areas. This signature is stronger in cisgender compared to transgender individuals given a sex/gender interaction. This signature may interact significantly with psychiatric disorders such as MDD (paper II).

Conclusion and outlook

The current thesis presents compelling evidence for the positive value of implementing ML models in psychiatry. It demonstrates that considering persons in their entirety by integrating a manifold of data modalities per subject along with stratifying models according to both biological sex and gender are important steps contributing to the ultimate goal of precision psychiatry: personalized care.

Further studies involving larger number of subjects and integrating data from different cohorts should be pursued aiming for robust and generalizable models in order to consolidate these results. Prospectively, we aim to extend the investigation on sexual dimorphism developing a multimodal integrative approach that combines neuroanatomical and phenotypical data in populations with psychotic and mood disorders leading to a better understanding of the etiopathology of these groups of psychiatric illnesses and therefore to an improvement of mental health care.

Final note: we advocate for the importance of sex/gender as a variable of interest in medical, but particularly in psychiatric research aiming to achieve personalized medicine. By no means this aims to strengthen beliefs of disparities between sexes or genders. Enough data is available that sustains the notion of sex-based differences in biological systems such as the brain. Thus, gathering and integrating this information into models that aid medical treatment decisions will ultimately favour individuals while also facilitating the optimization of clinical processes.

Zusammenfassung

Support Vector Machines (SVMs) spielen eine transformative Rolle bei der Weiterentwicklung der Präzisionspsychiatrie. Als eine leistungsstarke überwachte maschinelle Lerntechnik haben SVMs die diagnostische Genauigkeit und die Vorhersage von Ergebnissen auf individueller Ebene erheblich verbessert. In dieser Dissertation wurden zwei Ansätze untersucht. Erstens evaluieren wir das Potenzial von SVMs bei der Integration verschiedener Datenquellen, um ein umfassendes Verständnis von psychischen Gesundheitszuständen zu bieten. Darüber hinaus untersuchten wir, wie SVMs dabei helfen können, biologische und psychosoziale Unterschiede wie das biologische Geschlecht (*sex*) und das soziale Geschlecht (*gender*) einzubeziehen, was letztlich zu maßgeschneiderten und wirksameren Behandlungen führen kann.

In der ersten Studie verwendeten wir SVMs, um die Vorteile der Integration multimodaler Neuroimaging-Daten bei der Unterscheidung zwischen gesunden Personen und solchen mit Schizophrenie aufzuzeigen. Unsere Hauptbefunde zeigten, dass die Kombination von strukturellen und funktionellen MRT-Daten im Ruhezustand die Klassifikationsgenauigkeit auf 75% Balanced Accuracy (BAC) verbesserte, verglichen mit etwa 70% BAC bei der Verwendung einzelner Modalitäten.

Darüber hinaus untersuchten wir die Interaktionen zwischen geschlechtsbezogenen Gehirnstrukturen, Geschlechtsidentität und affektiven Störungen unter Verwendung eines maschinellen Lernrahmens. SVMs wurden angewendet, um das biologische Geschlecht (weiblich vs. männlich) anhand neuroanatomischer Daten von sowohl cis- als auch transgender Personen zu klassifizieren. Das Modell erreichte eine BAC von 82% für Cisgender-Proben, während es für Transgender-Proben eine BAC von 67% erzielte. Bei Anwendung auf Personen mit Major Depression sank die Leistung auf 75% bzw. 55% BAC. Diese Ergebnisse deuten darauf hin, dass es einen starken Einfluss des biologischen Geschlechts auf die Gehirnstruktur gibt, der in bedeutendem Maße mit der Geschlechtsidentität und psychiatrischen Phänotypen interagiert und so eine Diffusion der neuroanatomischen Signatur des biologischen Geschlechts verursacht.

Diese Analysen unterstreichen das Potenzial des maschinellen Lernens, insbesondere von SVMs, in Kombination mit einer rigorosen Charakterisierung von Individuen basierend auf biologischem Geschlecht/Gender und multimodaler Datenintegration, um die Präzisionspsychiatrie voranzutreiben. Durch die Erleichterung effektiverer diagnostischer, präventiver und therapeutischer Strategien verspricht dieser Ansatz, die Ergebnisse bei Personen mit psychischen Störungen zu verbessern.

Schlüsselwörter: maschinelles Lernen, SVM (Support Vector Machine), multimodale Datenanalyse, Neuroimaging, biologisches Geschlecht, Gender, soziale Geschlecht

Abstract

Support Vector Machines (SVMs) have a transformative role in advancing precision psychiatry. As a powerful supervised machine learning technique, SVMs have significantly enhanced diagnostic accuracy and outcome prediction on an individual level. In this thesis, two approaches were studied. First, we evaluate the potential of SVMs when integrating diverse data sources to provide a comprehensive understanding of mental health conditions. Moreover, we evaluated how SVMs can aid in acknowledging biological and psychosocial differences such as a person's sex and gender which ultimately help the tailoring of more effective treatments.

In the first study, we employed SVMs to showcase the benefits of integrating multimodal neuroimaging data for distinguishing healthy individuals from those with schizophrenia. Our main findings demonstrated that combining structural and resting-state functional MRI data improved classification accuracy to 75% Balanced Accuracy (BAC), compared to approximately 70% BAC using single modalities.

Furthermore, we investigated the interactions between sex-related brain structures, gender identity and mood disorders using a machine learning framework. SVMs were applied to classify biological sex (female vs. male) using neuroanatomical data from both cisgender and transgender individuals. The model achieved 82% BAC for cisgender samples, whereas for transgender samples, it achieved 67% BAC. When applied to individuals with major depression, performance decreased to 75% and 55% BAC, respectively. These results suggests that the existence of a strong imprint of sex on brain structure which interacts majorly with gender identity and psychiatric phenotypes, causing a blurring in the biological sex neuroanatomical signature.

These analyses underscore the potential of machine learning, particularly SVMs, coupled with rigorous characterization of individuals based on sex/gender and multimodal data integration, to advance precision psychiatry. By facilitating more effective diagnostic, preventive, and treatment strategies, this approach holds promise for improving outcomes in individuals with mental disorders.

Keywords: machine learning, SVM, multimodal data analysis, neuroimaging, sex, gender

Paper I

Cabral, C., Kambeitz-Ilankovic, L., Kambeitz, J., Calhoun, V. D., Dwyer, D. B., Von Saldern, S., **Urquijo MF.,** Falkai P. & Koutsouleris, N. (2016). **Classifying schizophrenia using multimodal multivariate pattern recognition analysis: evaluating the impact of individual clinical profiles on the neurodiagnostic performance**. *Schizophrenia bulletin*, *42*(suppl_1), S110-S117. https://doi.org/10.1093/schbul/sbw053

https://academic.oup.com/schizophreniabulletin/article/42/suppl_1/S110/2413943

Paper II

Baldinger-Melich, P.*, **Urquijo Castro, M. F.*,** Seiger, R., Ruef, A., Dwyer, D. B., Kranz, G. S., Klöbl, M., Kambeitz, J., Kaufmann, U., Windischberger, C., Kasper, S., Falkai, P., Lanzenberger, R. & Koutsouleris, N. (2020). **Sex matters: a multivariate pattern analysis of sex-and gender-related neuroanatomical differences in cis-and transgender individuals using structural magnetic resonance imaging.** *Cerebral Cortex*, 30(3), 1345-1356. https://doi.org/10.1093/cercor/bhz170
*Authors contributed equally

https://academic.oup.com/cercor/article/30/3/1345/5542405

References

Allebone, J., Kanaan, R., & Wilson, S. J. (2018). Systematic review of structural and functional brain alterations in psychosis of epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, *89*(6), 611-617.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

American Psychiatric Association (APA). (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist*, *70*(9), 832-864.

Argyelan, M., Gallego, J. A., Robinson, D. G., Ikuta, T., Sarpal, D., John, M., Kingsley, P. B., Kane, J., Malhotra, A. K., & Szeszko, P. R. (2015). Abnormal resting state FMRI activity predicts processing speed deficits in first-episode psychosis. *Neuropsychopharmacology*, *40*(7), 1631-1639.

Bouman, W. P., & Arcelus, J. (2017). *The transgender handbook: A guide for transgender people, their families and professionals*.

Cahill, L. (2014). Equal≠ the same: sex differences in the human brain. Cerebrum: the Dana forum on brain science,

Cahill, L. (2017). An issue whose time has come. *Journal of neuroscience research*, *95*(1-2), 12-13.

Calhoun, V. D., & Sui, J. (2016). Multimodal fusion of brain imaging data: a key to finding the missing link (s) in complex mental illness. *Biological psychiatry: cognitive neuroscience and neuroimaging*, *1*(3), 230-244.

Cha, J., Ide, J. S., Bowman, F. D., Simpson, H. B., Posner, J., & Steinglass, J. E. (2016). Abnormal reward circuitry in anorexia nervosa: A longitudinal, multimodal MRI study. *Hum Brain Mapp*, *37*(11), 3835-3846.

Chekroud, A. M., Ward, E. J., Rosenberg, M. D., & Holmes, A. J. (2016). Patterns in the human brain mosaic discriminate males from females. *Proceedings of the National Academy of Sciences*, *113*(14), E1968-E1968.

Chen, J., Dong, G., Song, L., Zhao, X., Cao, J., Luo, X., Feng, J., & Zhao, X.-M. (2021). Integration of multimodal data for deciphering brain disorders. *Annual Review of Biomedical Data Science*, *4*, 43-56.

Clayton, J. A., & Collins, F. S. (2014). Policy: NIH to balance sex in cell and animal studies. *Nature News*, *509*(7500), 282.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273-297.

Davy, Z., & Toze, M. (2018). What Is Gender Dysphoria? A Critical Systematic Narrative Review. *Transgend Health*, *3*(1), 159-169.

De Castro, P., Heidari, S., & Babor, T. F. (2016). Sex and gender equity in research (SAGER): reporting guidelines as a framework of innovation for an equitable approach to gender medicine. *Annali dell'Istituto superiore di sanita*, *52*(2), 154-157.

de Freitas, L. D., Léda-Rêgo, G., Bezerra-Filho, S., & Miranda-Scippa, Â. (2020). Psychiatric disorders in individuals diagnosed with gender dysphoria: A systematic review. *Psychiatry and clinical neurosciences*, *74*(2), 99-104.

Dekaban, A. S., & Sadowsky, D. (1978). Changes in brain weights during the span of human life: relation of brain weights to body heights and body weights. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *4*(4), 345-356.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annu Rev Clin Psychol*, *14*, 91-118.

Egloff, L., Lenz, C., Studerus, E., Harrisberger, F., Smieskova, R., Schmidt, A., Huber, C., Simon, A., Lang, U. E., Riecher-Rössler, A., & Borgwardt, S. (2018). Sexually dimorphic subcortical brain volumes in emerging psychosis. *Schizophrenia research*.

Eid, R. S., Gobinath, A. R., & Galea, L. A. M. (2019). Sex differences in depression: Insights from clinical and preclinical studies. *Progress in neurobiology*, *176*, 86-102.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118.

Feder, S., Isserlin, L., Seale, E., Hammond, N., & Norris, M. L. (2017). Exploring the association between eating disorders and gender dysphoria in youth. *Eating Disorders*, *25*(4), 310-317.

Fernandez-Guasti, A., Fiedler, J., Herrera, L., & Handa, R. (2012). Sex, stress, and mood disorders: at the intersection of adrenal and gonadal hormones. *Hormone and Metabolic Research*, *44*(08), 607-618.

Feusner, J. D., Lidström, A., Moody, T. D., Dhejne, C., Bookheimer, S. Y., & Savic, I. (2017). Intrinsic network connectivity and own body perception in gender dysphoria. *Brain Imaging and Behavior*, *11*, 964-976.

Flint, C., Förster, K., Koser, S. A., Konrad, C., Zwitserlood, P., Berger, K., Hermesdorf, M., Kircher, T., Nenadic, I., & Krug, A. (2020). Biological sex classification with structural MRI data shows increased misclassification in transgender women. *Neuropsychopharmacology*, *45*(10), 1758-1765.

Fusar-Poli, P., & Meyer-Lindenberg, A. (2016). Forty years of structural imaging in psychosis: promises and truth. *Acta Psychiatrica Scandinavica*, *134*(3), 207-224.

García-Acero, M., Moreno, O., Suárez, F., & Rojas, A. (2020). Disorders of Sexual Development: Current Status and Progress in the Diagnostic Approach. *Curr Urol*, *13*(4), 169-178.

Garg, G., Elshimy, G., & Marwaha, R. (2018). Gender dysphoria.

Gonzalez, C. A., Gallego, J. D., & Bockting, W. O. (2017). Demographic characteristics, components of sexuality and gender, and minority stress and their associations to excessive alcohol, cannabis, and illicit (noncannabis) drug use among a large sample of transgender people in the United States. *The journal of primary prevention*, *38*(4), 419-445.

Guillamon, A., Junque, C., & Gómez-Gil, E. (2016). A Review of the Status of Brain Structure Research in Transsexualism [journal article]. *Archives of Sexual Behavior*, *45*(7), 1615-1648.

Healy, D. (2015). Serotonin and depression. In (Vol. 350): British Medical Journal Publishing Group.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, *31*(8), 651-666.

Joel, D., Berman, Z., Tavor, I., Wexler, N., Gaber, O., Stein, Y., Shefi, N., Pool, J., Urchs, S., & Margulies, D. S. (2015). Sex beyond the genitalia: The human brain mosaic. *Proceedings of the National Academy of Sciences*, *112*(50), 15468-15473.

Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., Popovic, D., Oeztuerk, O., Haas, S. S., & Weiske, J. (2020). Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry*.

Kurth, F., Gaser, C., & Luders, E. (2021). Development of sex differences in the human brain. *Cognitive Neuroscience*, *12*(3-4), 155-162.

Kurth, F., Gaser, C., Sánchez, F. J., & Luders, E. (2022). Brain sex in transgender women is shifted towards gender identity. *Journal of Clinical Medicine*, *11*(6), 1582.

Li, H., Cui, L., Cao, L., Zhang, Y., Liu, Y., Deng, W., & Zhou, W. (2020). Identification of bipolar disorder using a combination of multimodality magnetic resonance imaging and machine learning techniques. *BMC psychiatry*, *20*, 1-12.

Li, R., Ma, X., Wang, G., Yang, J., & Wang, C. (2016). Why sex differences in schizophrenia? *Journal of translational neuroscience*, *1*(1), 37-42.

Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Jancke, L., Steinmetz, H., & Toga, A. W. (2004). Gender differences in cortical complexity. *Nature neuroscience*, *7*(8), 799-800.

Luders, E., Sanchez, F. J., Gaser, C., Toga, A. W., Narr, K. L., Hamilton, L. S., & Vilain, E. (2009). Regional gray matter variation in male-to-female transsexualism. *Neuroimage*, *46*(4), 904-907.

Mechelli, A., & Vieira, S. (2020). From models to tools: clinical translation of machine learning studies in psychosis. *npj Schizophrenia*, *6*(1), 4.

Mueller, S., Keeser, D., Samson, A. C., Kirsch, V., Blautzik, J., Grothe, M., Erat, O., Hegenloh, M., Coates, U., & Reiser, M. F. (2013). Convergent findings of altered functional and structural brain connectivity in individuals with high functioning autism: a multimodal MRI study. *PloS one*, *8*(6), e67329.

Müller, M. B., & Holsboer, F. (2005). Hormones, stress and depression. In *Hormones and the Brain* (pp. 227-236). Springer

Nolan, I. T., Kuhner, C. J., & Dy, G. W. (2019). Demographic and temporal trends in transgender identities and gender confirming surgery. *Translational andrology and urology*, *8*(3), 184.

O'Neill, A., Mechelli, A., & Bhattacharyya, S. (2019). Dysconnectivity of large-scale functional networks in early psychosis: a meta-analysis. *Schizophrenia Bulletin*, *45*(3), 579-590.

Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1140-1152.

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *Machine learning* (pp. 101-121). Elsevier

Plitman, E., Patel, R., & Chakravarty, M. M. (2020). Seeing the bigger picture: multimodal neuroimaging to investigate neuropsychiatric illnesses. *J Psychiatry Neurosci*, *45*(3), 147.

Porter, A., Fei, S., Damme, K. S., Nusslock, R., Gratton, C., & Mittal, V. A. (2023). A meta-analysis and systematic review of single vs. multimodal neuroimaging techniques in the classification of psychosis. *Molecular psychiatry*, *28*(8), 3278-3292.

Rashid, B., & Calhoun, V. (2020). Towards a brain-based predictome of mental illness. *Hum Brain Mapp*, *41*(12), 3468-3535.

Reed, M. B., Handschuh, P. A., Klöbl, M., Konadu, M. E., Kaufmann, U., Hahn, A., Kranz, G. S., Spies, M., & Lanzenberger, R. (2023). The influence of sex steroid treatment on insular connectivity in gender dysphoria. *Psychoneuroendocrinology*, *155*, 106336.

Riecher-Rossler, A., Butler, S., & Kulkarni, J. (2018). Sex and gender differences in schizophrenic psychoses-a critical review. *Arch Womens Ment Health*.

Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., Reus, L. M., Alloza, C., Harris, M. A., Alderson, H. L., Hunter, S., Neilson, E., Liewald, D. C. M., Auyeung, B., Whalley, H. C., Lawrie, S. M., Gale, C. R., Bastin, M. E., McIntosh, A. M., & Deary, I. J. (2018). Sex Differences in the Adult Human Brain: Evidence from 5216 UK Biobank Participants. *Cerebral cortex (New York, N.Y. : 1991)*, *28*(8), 2959-2975.

Rosen, M., Betz, L. T., Schultze-Lutter, F., Chisholm, K., Haidl, T. K., Kambeitz-Ilankovic, L., Bertolino, A., Borgwardt, S., Brambilla, P., & Lencer, R. (2021). Towards clinical application of prediction models for transition to psychosis: a systematic review and external validation study in the PRONIA sample. *Neuroscience & Biobehavioral Reviews*, *125*, 478-492.

Rosenblatt, J. D. (2016). Multivariate revisit to "sex beyond the genitalia". *Proceedings of the National Academy of Sciences*, *113*(14), E1966-E1967.

Rotondi, N. K., Bauer, G. R., Travers, R., Travers, A., Scanlon, K., & Kaay, M. (2012). Depression in male-to-female transgender Ontarians: Results from the Trans PULSE Project. *Canadian Journal of Community Mental Health*, *30*(2), 113-133.

Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M.-C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews*, *39*, 34-50.

Rutledge, R. B., Chekroud, A. M., & Huys, Q. J. (2019). Machine learning and big data in psychiatry: toward clinical applications. *Current opinion in neurobiology*, *55*, 152-159.

Salvador, R., Canales-Rodríguez, E., Guerrero-Pedraza, A., Sarró, S., Tordesillas-Gutiérrez, D., Maristany, T., Crespo-Facorro, B., McKenna, P., & Pomarol-Clotet, E. (2019). Multimodal integration of brain images for MRI-based diagnosis in schizophrenia. *Frontiers in neuroscience*, *13*, 1203.

Savic, I., Garcia-Falgueras, A., & Swaab, D. F. (2010). Sexual differentiation of the human brain in relation to gender identity and sexual orientation. *Progress in brain research*, *186*, 41-62.

Sax, L. (2002). How common is intersex? a response to Anne Fausto-Sterling. *J Sex Res*, *39*(3), 174-178.

Schaffer, C. (1993). Selecting a classification method by cross-validation. *Machine learning*, *13*(1), 135-143.

Schuch, J. J. J., Roest, A. M., Nolen, W. A., Penninx, B. W. J. H., & de Jonge, P. (2014). Gender differences in major depressive disorder: Results from the Netherlands study of depression and anxiety. *Journal of affective disorders*, *156*, 156-163.

Schultz, C. C., Fusar-Poli, P., Wagner, G., Koch, K., Schachtzabel, C., Gruber, O., Sauer, H., & Schlösser, R. G. (2012). Multimodal functional and structural imaging investigations in psychosis research. *European archives of psychiatry and clinical neuroscience*, *262*, 97-106.

Spicer, J., & Sanborn, A. N. (2019). What does the mind learn? A comparison of human and machine learning representations. *Current opinion in neurobiology*, *55*, 97-102.

Stowell, J. T., Jha, P., Martinez-Jorge, J., Middlebrooks, E. H., Broderick, D. F., & Bhatt, A. A. (2022). Neuroradiology in transgender care: facial feminization, laryngeal surgery, and beyond. *RadioGraphics*, *42*(1), 233-249.

Thorne, N., Yip, A. K.-T., Bouman, W. P., Marshall, E., & Arcelus, J. (2020). The terminology of identities between, outside and beyond the gender binary–A systematic review. *Non-Binary and Genderqueer Genders*, 20-36.

Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.

Weiler, M., Agosta, F., Canu, E., Copetti, M., Magnani, G., Marcone, A., Pagani, E., Balthazar, M. L. F., Comi, G., & Falini, A. (2015). Following the spreading of brain structural changes in Alzheimer's disease: a longitudinal, multimodal MRI study. *Journal of Alzheimer's Disease*, *47*(4), 995-1007.

Witcomb, G. L., Bouman, W. P., Claes, L., Brewin, N., Crawford, J. R., & Arcelus, J. (2018). Levels of depression in transgender people and its predictors: Results of a large matched control study with transgender people accessing clinical services. *Journal of affective disorders*, *235*, 308-315.

Yu, K.-H., Zhang, C., Berry, G. J., Altman, R. B., Ré, C., Rubin, D. L., & Snyder, M. (2016). Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature communications*, *7*(1), 12474.

Zhang, Q., Goodman, M., Adams, N., Corneil, T., Hashemi, L., Kreukels, B., Motmans, J., Snyder, R., & Coleman, E. (2020). Epidemiological considerations in transgender health: A systematic review with focus on higher quality data. *Int J Transgend Health*, *21*(2), 125-137.

Zucker, K. J. (2017). Epidemiology of gender dysphoria and transgender identity. *Sexual health*, *14*(5), 404-411.

Acknowledgement

I would first like to thank my supervisor Prof. Dr. Nikolaos Koutsouleris for accompanying me along the way. Your expertise in machine learning and your insightful feedback motivated me and pushed me to this point.

I would also like to acknowledge my colleagues and friends from the clinical team at the psychiatric clinic of the LMU for the wonderful moments that made us grow and learn through hard and easier times. Working and collaborating with you is what I call teamwork.

I want to thank Dr. Pia Baldinger. It was amazing to experience how scientific research can develop into a friendship.

Thanks to Dr. Lana Kambeitz-Ilankovic for her valuable guidance throughout my studies. You provided me with the tools that I needed to choose the right direction and grow as a person and as a scientist.

To the participants that provided the data for these analyses, also a big thank you. I hope these results will help you get access to better healthcare.

To the coauthors of the publications, thank you for this wonderful opportunity of working together.

More than everything, I would like to thank both my Colombian and German families for their support since the day I decided to pursue this goal.

Friends, I know you know who you are. Thanks for making this road so much easier. Just knowing that you were there gave me the motivation to move forward even throughout the hardest times of this creative process. Special thanks to those who took their time to provide feedback to this work: Mario, Kate, Nora, Memo, Tate, Rachele. You are the best!

Finally, I could not have completed this dissertation without the support of my endlessly beloved husband, Anibal, who is my coach, my mentor, my friend, and my persona.

I dedicate this work to Magdalena and Alejandro. Hijos, if you want to pursue something, just go for it, I'll be here to support you. Mami, papi, parece que lo logré.