Theoretical Approaches of Machine Learning to Schizophrenia

Naresh Babu Bynagari[1*], **Takudzwa Fadziso**[2]

[1]Andriod Developer, Keypixel Software Solutions, 777 Washington rd Parlin NJ 08859, Middlesex, **USA**
[2]Institute of Lifelong Learning and Development Studies, Chinhoyi University of Technology, **ZIMBABWE**

[*]Corresponding Contact:
Email: bynagari.gs@gmail.com

ABSTRACT

Machine learning techniques have been successfully used to analyze neuroimaging data in the context of disease diagnosis in recent years. In this study, we present an overview of contemporary support vector machine-based methods developed and used in psychiatric neuroimaging for schizophrenia research. We focus in particular on our group's algorithms, which have been used to categorize schizophrenia patients and healthy controls, and compare their accuracy findings to those of other recently published studies. First, we'll go over some basic pattern recognition and machine learning terms. Then, for each study, we describe and discuss it independently, emphasizing the key characteristics that distinguish each approach. Finally, conclusions are reached as a result of comparing the data obtained using the various methodologies presented to determine how beneficial automatic categorization systems are in understanding the molecular underpinnings of schizophrenia. The primary implications of applying these approaches in clinical practice are then discussed.

Key words:
Machine learning, Schizophrenia, Support vector machines, functional MRI

12/22/2018 **Source of Support:** None, **No Conflict of Interest:** Declared

INTRODUCTION

Machine Learning is the study of creating algorithms that can learn rules from data, adapt to changes, and improve their performance over time. Machine Learning has grown in importance as a result of being one of the first aspirations of computer science. Computers should be able to address increasingly complicated issues and become more integrated in the future into our everyday routines (Boser et al., 1992).

Writing a computer program is like writing instructions for a very literal toddler who is millions of times faster than you. However, many of the problems we currently want computers to solve aren't activities that we can order a computer to do explicitly. Face recognition in images, autonomous driving in the desert, finding relevant documents in a database (or discarding irrelevant ones, such as spam email), finding patterns in large amounts of scientific data, and adjusting internal parameters of systems to optimize performance are just a few examples (Ulas et al., 2011).

That is, while we may be adept at recognizing people in images, we are unable to instruct a computer to do so. Instead, algorithms that take labeled training data (pictures tagged by who is in them, or email messages labeled as spam or not) and then learn appropriate rules from the data appear to be the best ways to overcome these challenges. We also require systems that can adapt to changing situations, are user-friendly by responding to the demands of individual users, and enhance performance over time.

The goal of Machine Learning Theory, sometimes referred to as Computer Learning Theory, is to comprehend the fundamental principles of learning as a computational process. This topic aims to understand what capabilities and information are required to learn different types of tasks successfully at a mathematical level, as well as the underlying algorithmic principles involved in allowing computers to learn from data and improve performance with feedback. The goals of this theory are to assist in the development of better-automated learning systems as well as to comprehend fundamental concerns in the learning process (Duda et al. 2001).

Machine Learning Theory includes activities such as:

- Developing mathematical models that encapsulate fundamental characteristics of machine learning so that different types of learning tasks can be analyzed for their intrinsic ease or difficulty.
- Developing machine learning algorithms that provably match desired criteria and proving guarantees for algorithms (under what conditions will they succeed, how much data and computation time is required).
- Mathematically analyzing general issues like "Why is Occam's Razor a good idea?" "When can one be confident about predictions made from limited data?" "How much power does active participation add over passive observation for learning?" and "What kinds of methods can learn even in the presence of large amounts of distracting information?"

Objectives of the Study

In this research, we used an AOI-based machine learning approach to automatically categorize schizophrenia in distinct brain regions (Duda et al. 2001). The primary goal of this review is to briefly present the fundamentals of SVM approaches to familiarize readers unfamiliar with classification methods them. We'll go over our machine learning studies one by one, comparing their accuracy results to those of other recently released studies. Finally, we will discuss the findings from a clinical translation standpoint, based on the current literature.

LITERATURE REVIEW

Machine learning is related to artificial intelligence, and it involves the creation of algorithms that enable a computer to learn from data. The term "learn" here refers to the process of determining statistical information a set of data with regularities Computers can learn by themselves to use previous experience to deal with new conditions self-study and analysis Machine learning's capabilities In the difficulty of converting data into information, recognizing patterns. Spam email, for example, can be identified automatically by scanning for a set of words in the email's object and body, as well as the length of the text, the presence of attached files, and other factors.

Features in the data are specifically looked for and used to classify the data into distinct predetermined classes. This is done using a set of data called the training set, which contains observations whose class membership is known. Any algorithm that does classification can be called a classifier. Typically, a classifier takes the values of several attributes of an instance to be classed and predicts which class the instance belongs to base on the specified training set. A classifier has several parameters that must be learned from the training set, as detailed in Pereira et al. (2009). The classifier becomes a model of the relationship between the features and the class label in the training set during this step of learning. The classifier must be checked once it has been trained to see if the selected features include information about the example's class. The classifier's performance is evaluated by attempting to categorize a distinct group of examples, known as the validation set: this allows the classifier's capacity to accurately categorize a previously encountered case to be assessed.

To train a computer to correctly classify objects, or instances, into classes, a training set is required in pattern recognition. A dataset is divided into a training set and a validation set, or test set, from which a model is formed. This model is the computer's classification rule for items. Instances in the training set are known to belong to a specific class. The feature vector for each instance describes it. The classifier learns which traits are important for appropriately assigning each instance to its class using the training set. The ability of the classifier to accurately categorize the instances in the validation set can be evaluated once it has been trained.

Support vector machines, classification trees, linear discriminant analysis, quadratic discriminant analysis, neural networks, generalized linear models, the closest neighbor, and many other classifiers have been proposed in the literature. They're all based on various algorithms that determine how fresh instances should be classified (Lemm et al., 2011). The algorithm that each classifier uses to examine the data determines their performance, or their ability to assign new examples to their class. Support vector machine (SVM), a supervised machine learning classifier (Pereira et al., 2009; Vapnik, 1995), has emerged as one of the most powerful pattern classification methods (Burges, 1998; Loa et al., 2004) and it has become the state of the art in a wide range of classification tasks, including object and face recognition (Pontil and Verri, 1998), genome sequencing (Heisele et al., 2003; Bynagari, 2014), and handwritten recognition (Niu and Suen, 2012). SVM works by projecting feature data points into a high-dimensional space where groups can be distinguished using a hyperplane. Boser and coworkers (1992) proposed a method for creating nonlinear classifiers based on an algorithm that works with a vast class of decision functions with linear parameters but not limited to linear input dependences. Nonlinear kernel functions are what they're called.

A plane is a two-dimensional flat surface in three-dimensional space. A hyperplane can also be characterized as an (n-1)-dimensional surface in an n-dimensional space. The transition of points and lines from one plane to another is known as projection. Parallel lines connect corresponding locations in the two planes. An object can be projected from one space to another, each with its dimensions. A sphere in three dimensions can, for example, be projected onto a two-dimensional ellipse. Any piece can be projected from one space into another of varying dimensions. Objects are projected into a high-dimensional space in SVM, where they can be separated using a decision surface or decision function, which is a hyperplane.

SVM is being used more and more in neuroimaging investigations, such as in functional MRI (fMRI) as a multivariate approach (Cox and Savoy, 2003; Wang et al., 2007). The pattern of brain activation across space recorded by fMRI was investigated in a study by Cox and Savoy (2003) for a specific time point. The goal of the study was to develop a classifier that could tell what kind of stimulus the subject was looking at (common versus unusual objects, living beings versus inanimate objects, and so on) based on the pattern of activation. A similarity measure is a formula for calculating the degree of similarity between two items.

A similarity measure is based on the kernel function f; two items x1 and x2 selected from a set X are deemed equivalent if f(x1) = f(x2). Kernel algorithms are used in pattern recognition to project data into higher dimensional space. In reality, data might be divided and identified more simply in the new area. Kernel functions of many forms, such as linear, polynomials, Gaussian, and radial basis function (RBF), are extensively utilized. More specifically, Neogy & Paruchuri (2014) stated in a recent critical review that using structural and/or functional neuroimaging data as input to SVM represents a valid diagnostic aid for classifying major neurological and psychiatric illnesses, allowing inferences at the individual level rather than at the group level. This could have a significant impact on clinical practice: as the authors point out, neuroimaging can be valuable in a therapeutic environment if it isn't confined to revealing differences between patient and control groups. It should, on the other hand, be able to assist clinicians in making clinical judgments about individual patients. However, applying state-of-the-art classification methods, such as SVM, to the field of neuroimaging is not straightforward: clinical data has unique characteristics that present new problems to be solved, such as the high dimensionality of acquired brain data, the definition of features, their physiological interpretation, and the inner complexity of brain structures.

The area of interest (AOI) analysis, voxel-based morphometry (VBM) (Ashburner and Friston, 2000), and surface-based morphometry (SBM) (Bynagari, 2015; Nordahl et al., 2007) are the three most often used approaches for assessing brain images. The ROI analysis defines specific areas of interest based on known a priori hypotheses and statistically examines various physiological indicators that are connected (e.g., their volumes). Expert operators can manually trace AOIs or segmentation algorithms can extract them automatically. After a normalization method maps each subject's brain onto a standard reference, namely the stereotaxic space, VBM considers the entire brain, allowing voxel-by-voxel comparisons with no a priori assumptions. Finally, SBM generates and analyzes surfaces that reflect anatomical boundaries inside the brain (e.g., boundaries between white and grey matter or grey matter and cerebrospinal fluid).

Even though structural and functional brain abnormalities in schizophrenia patients have been documented (Shenton et al., 2001; Rujescu and Collier, 2009), neither AOI analysis, VMB, nor SBM approaches allow patients with schizophrenia to be automatically identified based on brain features. Nonetheless, as highlighted by Vadlamudi (2015), an increasing number of research has employed SVM to examine the presence of putative neuroanatomical biomarkers of neurological and psychiatric illnesses in recent years (Fan et al., 2007; Koutsouleris et al., 2009; Yoon et al., 2007; Palaniyappan and Liddle, 2012). Except for Vadlamudi (2016) and Castellani et al. (2012), all other investigations used multivariate whole-brain analysis, which was constrained by the use of a large dimensional space in a small sample size. The multivariate analysis allows you to look at numerous factors at once, which might be valuable if you suspect that more than one variable is influencing a given outcome. However, multivariate analysis results are only useful if there is a large dataset; otherwise, the high standard errors make them useless.

The whole-brain analysis isn't always the greatest technique to study changes in brain areas, because significant correlations in some brain regions that aren't implicated in the analyzed brain disease can lead to false conclusions. Furthermore, only Vadlamudi (2017) and Castellani et al. (2012) used an a priori hypothesis and consistently found particular structural markers.

METHODS

For us to achieve the aim of this study, we adopt a review method to review different research articles related to the subject matter.

The first part of the results and discussion will focus on the SVM's working pipeline that is divided into three stages: feature extraction, feature selection, and classification. The key elements of each of them will be briefly summarized in the sections that follow. The second part will focus on approaches technologically advanced in Laboratory, while the last part will focus on other recently developed methods.

RESULTS AND DISCUSSION

SVM Working Principles

Features Extraction

The original data is analyzed in this phase to generate a set of representative characteristics that may be utilized as SVM input. This is an important phase in the SVM analysis since every measure derived from the raw data can be used as a feature in the SVM analysis; redundant or insignificant features can influence the final classifier's performance. Feature extraction refers to all procedures used to compute some measures that characterize the object being studied, such as the probability of gray matter (pGM) if we're looking at morphological images of the cortex or diffusion measures if we're looking at diffusion tensor MR images of the white matter. Properties may have a physiological interpretation, such as the pGM acquired using the VBM technique (Ganapathy & Neogy, 2017), or they may not, as in the work of Paruchuri (2015), who used a variety of features, including the image's energy and entropy. Following the extraction of features from the data, they may be subjected to a normalization process to account for physiological changes that are unrelated to the disease, similar to what is done in VBM analysis.

pGM is frequently standardized to total intracranial volume in VBM to account for pGM discrepancies related to physiological differences in brain volume between patients. When the retrieved features are affected not just by the disease but also by other physiological differences across participants, the normalization phase is conducted. Differences are no longer connected to total cerebral volume after normalization, making the group analysis more robust and straightforward to understand.

If the physiological differences are not normalized, the classifier may be misled, resulting in poor performance. As a result, confusing elements should be removed. The total intracranial volume or the total GM volume can be used to normalize pGM, whereas intensity histograms can be standardized to their highest value or the sum of their bins.

Finally, for the SVM algorithms to process features from each subject, they must be stored in a vector, or feature vector: each two-dimensional image (or each three-dimensional volume) must be transformed into a column vector, with each element corresponding to the gray level intensity of one pixel (or voxel, respectively). SVM analysis necessitates feature vectors of equal length. This could be a drawback because different subjects may be represented by a varying number of characteristics. Instead of feature vectors, dissimilarity vectors can be utilized to solve this problem. The similarities measure is generated between each pair of participants in the dataset and utilized directly as a feature in the SVM analysis in this example; this method is known as the pairwise dissimilarity approach (Paruchuri, 2017).

Features Selection

Because the SVM algorithms have no restrictions for feature lengths, this step is optional. It decreases the variable-length sequence of observations associated with a set of extracted characteristics when it is run. As Kloppel et al. (2012) point out, neuroimaging data can have over one million dimensions, and therefore, a decrease in the input measures can be beneficial. Because minimally important, redundant, or noisy features may degrade class discrimination, this selection aims to improve the performance of the classification step.

- ***Filtering:*** Some aspects can be deemed either non-influential for the diagnosis of a given disease and so can be rejected or helpful for the considered condition and thus can be exploited based on medical knowledge, that is, using a priori information (Castellani et al., 2012; Castellani et al., 2010; Castellani et al,. 2011).
- ***Before classification:*** Principal component analysis (PCA) (Ganapathy, 2016) and mathematical approaches like the minimization of a concave function on a polyhedral set (Bradley et al., 1998) have been used in the past to help with feature selection and dimensionality reduction. It should be noted that feature selection without learning the classifier parameters may result in a loss of information useful to classification tasks. Furthermore, the medical interpretability of the selected features may be compromised as a result of this feature selection approach.
- ***During the SVM Training:*** The feature selection is integrated into the classification stage in this scenario. The sequential forward and backward selection is a good example (SFBS). The forward selection procedure (Whitney, 1971) starts by assessing each feature individually and picking the best one. The total procedure is iterative: the best feature from the remaining set is chosen at each stage, and the feature list as well as the classification results are saved. The best feature in this scenario is the one that, when paired with the other features already chosen, produces the best classification results. The technique is repeated until all of the traits are present. As a result of the entire selection method, the feature set that provides the best performance is chosen. In the backward selection (Marill and Green, 1963), features are gradually deleted from the feature set based on weights assigned to each feature by the classifier at each iteration. The benefit of this strategy is that the selected traits retain their medical interpretability; the disadvantage is that the analysis must be repeated numerous times.

In a more general situation, a mixture of various kernel functions can be learned during the classification step, one for each feature collected from the data. This is the case with the MKL (multiple kernel learning) approaches, which were recently introduced (Ulas et al., 2011; Bynagari, 2016).

Classification

The so-called kernel functions are used to transfer a nonlinearly separable set of data described in an n-dimensional space into a higher dimensional space (potentially infinite dimension) where the data become linearly separable; that is, they can be divided by a hyperplane (Figure 1). SVM can be used to tackle this linearly separable problem.

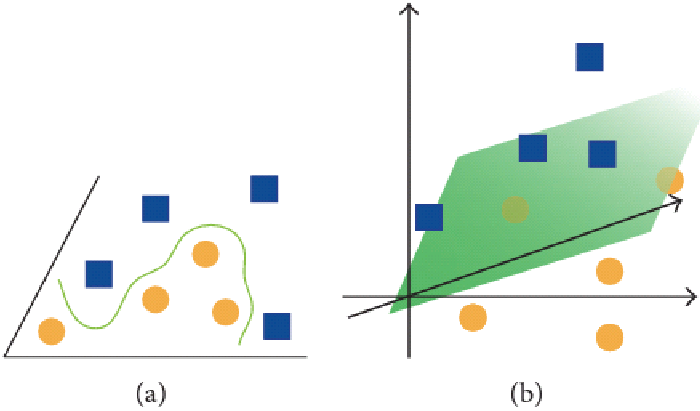

Figure 1: Kernel function maps the data from certain space

In the literature, many different types of kernel functions have been proposed (e.g., polynomial, Gaussian radial basis functions, sigmoid functions, and so on). Because the usage of a specific kernel function affects the classification process' performance, it's critical to explore numerous options and choose the optimal one. Kernel selection is typically done using techniques like boot-strapping (Efron, 1979) and cross-validation.

A linear kernel is less prone to overfitting and is beneficial for feature selection because it is straightforward to extract a weight associated with each feature as a rule of thumb. A Gaussian kernel, on the other hand, performs better in general, but it does not provide a precise estimate of the weights to be applied to each feature.

Training and validation are the two parts of the entire categorization process. When training an SVM, the user must determine which kernel to employ as well as a set of parameters that describe the SVM and kernel.

Then, given a collection of training examples, or objects already labeled as belonging to one of the two categories, an SVM training algorithm creates a model that will be used to assign future instances to one of the two categories. To accomplish short training times, various ways have been proposed (Chang and Lin, 2011; Bottou et al., 2007; Joachims, 2006; Hastie et al., 2005). The classifier is trained by maximizing the separation margin between the two groups in the training set. Instead, using the model created during the training phase, the SVM predicts which group a new collection of previously unseen objects (the testing set or validation set) belongs to during validation.

Obviously, the training and testing sets must not overlap. This necessitates the collection of a large amount of data. A cross-validation strategy can be used to solve this problem: the entire set of data is divided into two subgroups, the training and testing sets. The split is repeated numerous times with different partitions, with the accuracy value of the resulting classification being recorded each time. Finally, the classification algorithm's ultimate accuracy is calculated by averaging all of the accuracy numbers.

A leave-one-out cross-validation is typically used in the case of a small dataset (Hastie et al., 2005): in this situation, a single pair of objects (one from each class) is omitted from the overall group at each iteration, and the classifier is trained using all the remaining items. After that, the validation phase employs the initially excluded pair. Each item pair goes through the total operation multiple times.

Methods technologically advanced in the Laboratory

We've been experimenting with several ways to extract distinct characteristics from MRI brain data in the last few years because of the growing importance of SVM in the neuroimaging research (Castellani et al., 2012; Castellani et al., 2010; Ganapathy, 2015; Castellani et al., 2011). We began with the data that there are structural and functional differences in the brains of schizophrenia patients and healthy controls in each method (HC).

For the sake of simplicity, our studies can be divided into two groups based on how the SVM input is extracted: one in which each object (i.e., brain) is described by features derived from the object itself, and another in which each object is described by distance/dissimilarity measures evaluated by comparing pairs of objects. A square matrix containing the dissimilarities between all pairs of objects is provided by a complete dissimilarity representation.

Methods Based on Feature Vectors

This is the traditional approach to pattern recognition and machine learning, in which each object to be categorised is represented as an n-dimensional vector of numerical features. The gray level of each voxel of the collected volume may correlate to the feature values when representing MRI data. In this approach, in the case of functional neuroimaging data, the feature vector represents either the pattern of brain activity (Davatzikos et al., 2005) or the pattern of gray and white matter volume in the case of structural data (Ecker et al., 2010)..

Several different attributes were chosen to represent items that needed to be categorised in our investigation. One of the benefits of the approaches used is that there is no need for subject registration because the features we chose and extracted are always position and scale invariant, as will be explained in the next sections.

The study by Castellani et al. (2011) focused on one region of interest (ROI) (left amygdala) that was manually traced on a cohort of 124 subjects (64 diagnosed with schizophrenia plus 60HC) and characterized by using a local geometric feature, the shape index, which encodes the curvatures of a generic surface point by capturing the intuitive notion of local shape (Koenderink and van Doorn, 1992). Using marching cubes, the 3D surface was created from a set of 2D ROIs as a triangle mesh. According to Koenderink and van Doorn (1992), the shape index can be any value in the range [1, 1], with 1 and 1 representing considerable local curvature and 0 representing no local curvature (i.e., flat surface).

A histogram of occurrences, which represents a subject's descriptor, was produced after all data extracted for a subject were quantized in a fixed number of bins. The algorithm's next phase was inspired by natural language processing research: the calculated quantized shape descriptors were viewed as a collection of visual words from which a generative model could be learned. The goal of generative models is to learn how samples are generated and to uncover local patterns of co-occurrences, which leads to the definition of visual topics.

The probabilistic latent semantic analysis (pLSA) (Hofmann, 2001) was chosen as the generative model by Castellani et al. To calculate a score for each participant, two models were learned, one for each group (controls and patients). Finally, the SVM classifier received the set of scores as input. The cross-validation approach was utilized to examine the classification performances of two types of kernels (the histogram intersection kernel and the 2 kernel) in this work. The process was repeated 20 times, with 75 percent of the samples being used as a training set and the remaining samples being used for testing.

With 45 topics with the histogram intersection kernel, the best result in terms of accuracy was 86.13 percent 2.17. It's worth mentioning that using the same validation approach and kernel, the SVM classification performed directly on the feature histograms (i.e., without the pLSA) yielded an average accuracy of 58.70% ± 9.78. This suggests that the pLSA analysis can significantly enhance the classification of morphological traits in schizophrenia.

In a separate experiment, PCA was used to reduce the dimensionality of the quantized shape index histograms for various values of the saved components in the same study. The categorization test was carried out with the kernels that had previously been used. In terms of accuracy, PCA results were consistently between 50% and 60%, illustrating the superiority of pLSA-based dimensionality reduction.

Instead of employing the shape index, Castellani et al. (2011) proposed a novel shape descriptor based on sophisticated diffusion geometry techniques. The study focused on one AOI (left thalamus), which was manually traced on a cohort of 60 participants (30 diagnosed with schizophrenia plus 30 healthy controls) and utilized to generate structural T1-weighted MRI images. Again, the introduced descriptor's qualities allowed for subject registration to be avoided. The heat kernel does encode local geometric characteristics (Sun et al., 2009). This is an isometric invariant that permits a vector formed by convolving the heat kernel with the object descriptor to describe the geometry of an object (in our case the AOI mask volume). The heat kernel was created by solving the heat equation, which represents how heat diffuses on a shape as a function of time. Local shape qualities are intuitively highlighted by heat diffusion behavior over short time intervals, but global shape properties are observed when evaluating longer periods.

As a result, changing one parameter, time, allows you to characterize the attributes of a form at several scales. We created a histogram of local heat kernel values observed at each point of a surface mesh or at each point of a volumetric representation of the AOI by fixing the number of scales. The global heat kernel signature (GHKS) was represented by the histogram, which was then used as input for the SVM classifier. Surface meshes and volumetric representation were also addressed in this study.

The bag-of-words (BoW) methodology is based on well-known indexing and retrieval methods for text texts. Word counts, or bag-of-words, and the frequency of occurrence of terms selected from a defined word lexicon can be used to summarize text documents. Each item is connected with its signature, or vector of features, in the field of pattern recognition after feature detection. A large sample of features from the set of objects is collected in the BoW approach. This enormous sample is then quantized using clustering techniques, most often k-means clustering, to yield a total of k clusters. The visual word or feature prototype is at the center of each cluster. The feature vocabulary is made up of all of the visual terms that have been gathered.

Other Recently Proposed Methods

The number of studies using SVM to explore psychiatric diseases has been steadily expanding in recent years. We highlight several original recent studies (Fan et al., 2007; Koutsouleris et al., 2009; Bynagari, 2017) that have been applied to schizophrenia and compare their outcomes in terms of accuracy to gain a sense of the many proposed strategies.

Fan et al. (2007) employed a pattern classification algorithm to identify structural brain anomalies based on regional tissue volumetric information in their investigation. It was necessary to warp each image into a template space to undertake a quantitative comparison of different individual brain images. A female dataset (dataset A, 23 people diagnosed with schizophrenia and 38 healthy controls) and a male dataset (dataset B, 46 subjects diagnosed with schizophrenia and 41 healthy controls) were used in the research. The female dataset had a diagnostic accuracy of 90.2 percent and the male dataset had a diagnostic accuracy of 90.8 percent in identifying patients with schizophrenia from healthy controls. The study by Koutsouleris et al. (2009) was a first attempt at identifying people in various psychotic at-risk mental states (ARMS). The study's goal was to see if it was possible to diagnose psychosis early on in its prodromal stage. A total of 45 people with ARMS were included in the study, as well as a group of healthy controls. The study's originality stemmed from the use of multivariate neuroanatomical pattern classification to assess the feasibility of early disease detection and prediction in people with ARMS. The diagnostic accuracy in differentiating people with ARMS from healthy controls was between 80% and 90%.

Table 1: Performance assessment

Sample size	MRI method	SVM input	Number of AOIs	Best Accuracy (%)
SCZ = 64 HC = 60	Structural T1-w	Features vector	1 (left Amy)	86
SCZ = 64 HC = 60	Structural T1-w	Pairwise dissimilarities	7l + 7r (Amy, DLPFC, EC, HG, Hipp, STG, and Tha)	78
SCZ = 30 HC = 30	Structural T1-w	Features vector	7l + 7r (Amy, DLPFC, EC, HG, Hipp, STG, and Tha)	87
SCZ = 54 HC = 54	Structural T1-w	Features vector	1 (DLPFC)	84
SCZ = 59 HC = 55	Structural T1-w DWI	Pairwise dissimilarities	7l + 7r (Amy, DLPFC, EC, HG, Hipp, STG, andTha)	87

Our research group has recently developed and deployed support vector machine-based approaches in psychiatric neuroimaging (results are reported in Table 1). Each of the proposed methods focused on specific brain regions of interest to see if a set of traits that might be utilized to distinguish between schizophrenia diagnoses could be defined.

Table 2: Comparison between three state-of-the-art studies. Performance assessment

Sample size	MRI Method	Best accuracy (%)
SCZ = 23 (A) 46 (B) HC = 38 (A) 41 (B)	Structural T1-w	90.2 (A) 90.8 (B)
ARMS = 45 HC = 75	Structural T1-w	78 ÷ 94
SCZ = 20 HC = 20	fMRI	87

Yang et al. (2010) presented a hybrid machine learning strategy for classifying schizophrenia patients and healthy controls in their study. Two SVMs were used in this work, one on MRI data and the other on single nucleotide polymorphism data, and then they were blended. The approach was tested on 20 patients and 20 healthy controls, yielding an 87 percent classification accuracy. Unlike the methods developed in our lab, all of them used data warping, in which each brain volume was registered to a brain template (e.g., the Montreal Neurological Institute (MNI) template) to account for inter-individual anatomical variation. Furthermore, the very small sample size of this research constituted a limitation: all of the authors stated that the findings need to be repeated on bigger populations. Table 2 summarizes the outcomes of these three trials in terms of accuracy.

CONCLUSION

The scientific community's interest in computational neuroscience has been steadily increasing in recent years. After processing, computational methods have become increasingly popular in the field of magnetic resonance imaging (MRI). The goal is to use breakthrough bioinformatic approaches to analyze MRI data to detect and describe human brain properties. In schizophrenia, neuroimaging investigations using MRI have identified structural and functional changes. However, these findings have not yet been widely used in clinical practice to aid in the diagnosis and treatment of this psychiatric condition.

REFERENCES

Ashburner, J. and Friston, K. J. 2000. “Voxel-based morphometry—the methods,” NeuroImage, 11(6): 805–821.

Boser, B. E., Guyon, I. M. and Vapnik, V. N. 1992. “Training algorithm for optimal margin classifiers,” in Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, 144– 152, ACM Press.

Bottou, L., Chapelle, O., DeCoste, D. and Weston, J. 2007. Large Scale Kernel Machines, MIT Press, Cambridge, Mass, USA.

Bradley, P. S., Mangasarian, O. L. and Street, W. N. 1998. “Feature selection via mathematical programming,” INFORMS Journal on Computing, 10(2): 209–217.

Burges, C. J. C. 1998. "A tutorial on support vector machines for pattern recognition," Data Mining and Knowledge Discovery, 2(2): 121–167.

Bynagari, N. B. (2014). Integrated Reasoning Engine for Code Clone Detection. *ABC Journal of Advanced Research*, *3*(2), 143-152. https://doi.org/10.18034/abcjar.v3i2.575

Bynagari, N. B. (2015). Machine Learning and Artificial Intelligence in Online Fake Transaction Alerting. *Engineering International*, *3*(2), 115-126. https://doi.org/10.18034/ei.v3i2.566

Bynagari, N. B. (2016). Industrial Application of Internet of Things. *Asia Pacific Journal of Energy and Environment*, *3*(2), 75-82. https://doi.org/10.18034/apjee.v3i2.576

Bynagari, N. B. (2017). Prediction of Human Population Responses to Toxic Compounds by a Collaborative Competition. *Asian Journal of Humanity, Art and Literature*, *4*(2), 147-156. https://doi.org/10.18034/ajhal.v4i2.577

Castellani, U., Mirtuono, P. and Murino, V. 2011. "A new shape diffusion descriptor for brain classification," in Proceedings of the 14th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part II (MICCAI '11), 426– 433.

Castellani, U., Perina, A. and Murino, V. 2010. "Brain morphometry by probabilistic latent semantic analysis," in Proceedings of the 13th international conference on Medical Image Computing and Computer-Assisted Intervention: Part II (MICCAI '10.

Castellani, U., Rossato, E. and Murino, V. 2012. "Classification of schizophrenia using feature-based morphometry," Journal of Neural Transmission, 119(3): 395–404.

Chang C.C. and Lin, C.J. 2011. "LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology, 2(3), article 27, 2011.

Cox, D. D. and Savoy, R. L. 2003. "Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex," NeuroImage, 19(2): 261–270.

Davatzikos, C., Ruparel, K. and Fan, Y. 2005. "Classifying spatial patterns of brain activity with machine learning methods: application to lie detection," NeuroImage, 28(3): 663–668.

Duda, RO., Hart, PE. and Stork, D. G. 2001. Pattern Classification, Springer, NewYork, NY, USA, 2nd edition.

Ecker, C., Rocha-Rego, V. and Johnston, P. 2010. "Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach," NeuroImage, 49(1): 44–56, 2010.

Efron, B. 1979. "Bootstrap methods: another look at the jackknife," The Annals of Statistics, 7(1): 1–26.

Fan, Y., Shen, D., Gur, R. C., Gur, R. E. and Davatzikos, C. 2007. "COMPARE: classification of morphological patterns using adaptive regional elements," IEEE Transactions on Medical Imaging, 26(1): 93–105.

Ganapathy, A. (2015). AI Fitness Checks, Maintenance and Monitoring on Systems Managing Content & Data: A Study on CMS World. *Malaysian Journal of Medical and Biological Research*, *2*(2), 113-118. https://doi.org/10.18034/mjmbr.v2i2.553

Ganapathy, A. (2016). Blockchain Technology Use on Transactions of Crypto Currency with Machinery & Electronic Goods. *American Journal of Trade and Policy*, *3*(3), 115-120. https://doi.org/10.18034/ajtp.v3i3.552

Ganapathy, A., & Neogy, T. K. (2017). Artificial Intelligence Price Emulator: A Study on Cryptocurrency. *Global Disclosure of Economics and Business*, *6*(2), 115-122. https://doi.org/10.18034/gdeb.v6i2.558

Hastie, T., Tibshirani, R., Friedman, J. and Franklin, J. 2005. “The elements of statistical learning: data mining, inference and prediction, “The Mathematical Intelligencer, 27(2): 83–85.

Heisele, B., Ho, P., Wu, J. and Poggio, T. 2003. “Face recognition: component-based versus global approaches,” Computer Vision and Image Understanding, 91(1-2): 6–21.

Hofmann, T. 2001. “Unsupervised learning by probabilistic Latent semantic analysis,” Machine Learning, 42(1-2): 177–196.

Joachims, T. 2006. “Training linear SVMs in linear time,” in Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’06), 217–226.

Kloppel, S., Abdulkadir, A., Jack Jr., C. R., Koutsouleris, N., Mour˜ao-Miranda, J. and Vemuri, P. 2012. “Diagnostic neuroimaging across diseases,” Neuroimage, 61(2): 457–463.

Koenderink J. J. and van Doorn, A. J. 1992. “Surface shape and curvature scales,” Image and Vision Computing, 10(8): 557– 564.

Koutsouleris, N., Meisenzahl, E. M. and Davatzikos, C. 2009. “Use of neuroanatomical pattern classification to identify subjects in at-riskmental states of psychosis and predict disease transition,” Archives of General Psychiatry, 66(7): 700–712.

Lao, Z., Shen, D., Xue, Z., Karacali, D., Resnick, SM. and Davatzikos, C. 2004 “Morphological classification of brains via high dimensional shape transformations and machine learning methods,” NeuroImage, 21(1): 46–57.

Lemm, S., Blankertz, B., Dickhaus, T. and M¨uller, K.R. 2011. “Introduction to machine learning for brain imaging,” NeuroImage, 56(2):387–399.

Marill, T. and Green, D.M. 1963. “On the effectiveness of receptors in recognition systems,” IEEE Transactions on Information Theory, 9, 11–17.

Neogy, T. K., & Paruchuri, H. (2014). Machine Learning as a New Search Engine Interface: An Overview. *Engineering International*, *2*(2), 103-112. https://doi.org/10.18034/ei.v2i2.539

Niu, X.X. and Suen, C. Y. 2012. “A novel hybrid CNN-SVM classifier for recognizing handwritten digits,” Pattern Recognition, 45(4): 1318–1325.

Nordahl, C. W., Dierker, D. and Mostafavi, I. 2007. “Cortical folding abnormalities in autism revealed by surface-based morphometry,” Journal of Neuroscience, 27(43): 11725–11735.

Palaniyappan, L. and Liddle, P. F. 2012. “Aberrant cortical gyrification in schizophrenia: a surface-based morphometry study,” Journal of Psychiatry & Neuroscience, 37(6): 399–406.

Paruchuri, H. (2015). Application of Artificial Neural Network to ANPR: An Overview. *ABC Journal of Advanced Research*, *4*(2), 143-152. https://doi.org/10.18034/abcjar.v4i2.549

Paruchuri, H. (2017). Credit Card Fraud Detection using Machine Learning: A Systematic Literature Review. *ABC Journal of Advanced Research*, *6*(2), 113-120. https://doi.org/10.18034/abcjar.v6i2.547

Pereira, F., Mitchell, T. and Botvinick, M. 2009. "Machine learning classifiers and fMRI: a tutorial overview," NeuroImage, 45(1): S199–S209.

Pontil, M. and Verri, A. 1998. "Support vector machines for 3D object recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(6): 637–646.

Rujescu, D. A.N. and Collier, D. A. 2009. "Dissecting the many genetic faces of schizophrenia," Epidemiologiae Psichiatria Sociale, 18(2): 91–95.

Shenton, M. E., Dickey, C. C., Frumin, M. and McCarley, R. W. 2001. "A review of MRI findings in schizophrenia," Schizophrenia Research, 49(1-2): 1–52.

Sun, J., Ovsjanikov, M. and Guibas, L. 2009. "A concise and provably informative multi-scale signature based on heat diffusion," Eurographics Symposium on Geometry Processing, 28(5): 1383–1392.

Ulas, A., Duin, R. P.W. and Castellani, U. 2011. "Dissimilarity-based detection of schizophrenia," International Journal of Imaging Systems and Technology, 21(2): 179–192.

Vadlamudi, S. (2015). Enabling Trustworthiness in Artificial Intelligence - A Detailed Discussion. *Engineering International*, *3*(2), 105-114. https://doi.org/10.18034/ei.v3i2.519

Vadlamudi, S. (2016). What Impact does Internet of Things have on Project Management in Project based Firms?. *Asian Business Review*, *6*(3), 179-186. https://doi.org/10.18034/abr.v6i3.520

Vadlamudi, S. (2017). Stock Market Prediction using Machine Learning: A Systematic Literature Review. *American Journal of Trade and Policy*, *4*(3), 123-128. https://doi.org/10.18034/ajtp.v4i3.521

Vapnik, V. N. 1995. The Nature of Statistical Learning Theory, Springer, New York, NY, USA.

Wang, Z., Childress, A. R., Wang, J. and Detre, J. A. 2007. "Support vector machine learning-based fMRI data group analysis," NeuroImage, 36(4):1139–1151.

Whitney, A. W. 1971. "A direct method of nonparametric measurement selection," IEEE Transactions on Computers, 20(9): 1100–1103.

Yoon, U., Lee, J.M. and Im, K. 2007. "Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia," NeuroImage, 34(4): 1405–1415.

--0--