Protecting fMRI data from unforeseen privacy attacks in a distributed machine learning environment

Shan Suthaharan, Michael Ellis, Naseeb Thapaliya, Lavanya Goluguri, Harini Suresh Booravalli
University of North Carolina at Greensboro
Greensboro, North Carolina, USA
{s_suthah,mtellis2,n_thapal,l_golugu,h_boorav}@uncg.edu

ABSTRACT

The availability of large number of correlated features of the brain regions (as voxels) in the functional magnetic resonance imaging (fMRI) data can help us develop predictive models and extract brain networks for neuroscience applications. On the other hand, the availability of such a large set of features that describe the communication between brain regions can make the fMRI data vulnerable to unforeseen privacy attacks, especially when the data is distributed and scattered over multiple computing platforms. The purpose of this paper is to present a computational technique that compresses the features by treating them as signals to eliminate hidden patterns – as a feature selection approach – while preserving the patterns that are relevant for making predictive models efficient. For this purpose, we used the concept of compressed sensing and compressed learning techniques, combined with two-state Markov chain to integrate confusion and diffusion. We have used transition probabilities of the fMRI signals to construct compressed sensing matrix, transform the signals for compressed learning, and construct privacy-preserving predictive model. We have then evaluated the proposed computational technique using Logistic Regression, Sparse Multinomial Logistic Regression, Naïve Bayes, and Artificial Neural Networks. Our finding is that the proposed computational learning technique helps improve the performance of the predictive models overall, while enhancing their privacy-preserving capabilities through compressed sensing and cryptographic properties.

CCS CONCEPTS

• **Computing methodologies** → **Markov decision processes**;
• **Security and privacy** → *Privacy protections.*

KEYWORDS

datasets, neural networks, gaze detection, text tagging

ACM Reference Format:
Shan Suthaharan, Michael Ellis, Naseeb Thapaliya, Lavanya Goluguri, Harini Suresh Booravalli. 2019. Protecting fMRI data from unforeseen privacy attacks in a distributed machine learning environment. In *AIDR '18: Artificial Intelligence for Data Discovery and Reuse, May 13–15, 2019, Pittsburgh, PA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/1122445.1122456

AIDR '19, May 13–15, 2019, Pittsburgh, PA

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The current advances in artificial intelligence and computing technology allows the development of efficient computational learning techniques that are capable of discovering complex structures and functionalities of brain networks by using fMRI data and machine learning algorithms [14]. The brain network is unique to an individual [7]; hence, the extraction of structural (revelation of region of interests) and functional brain networks (communication between them) can disclose individual's brain maturity (e.g., individual thoughts and opinions) through the identification of communicating brain regions, when interpreting stimulus. Therefore, the accurate discovery of such brain networks can also help the unauthorized people (e.g., attackers) to infer privacy information of an individual. As a result, it is important for us to focus on privacy aspects of the neuroscience research, while developing efficient predictive models, using fMRI data with associated stimulus.

In the past couple of decades, a significant increase in neuroscience research using fMRI data analytics, and statistical or machine learning techniques can be observed. For example, in 2004, Mitchell et. al. [11] used machine learning to study classifiers that mainly address the problems of understanding the cognitive states of human subjects. In 2010, Ryali et. al. [16] studied the dimensionality problems in pattern recognition techniques used in fMRI data, and proposed a logistic regression-based method that incorporates both L1 and L2 regularization to extract distinguishable brain's region of interests. Subsequently in 2011, Naselaris et. al. [12] studied voxel-based encoding models in detail and recognized the inherent connections between the voxel-based encoding and decoding models using a linearized feature space, when the information is registered in brain's region of interests. In 2012, Cabral et. al. [2] studied Gaussian Naïve Bayes and k-NN classifiers to decode the brain's cognitive states under the constraints of high dimensionality and low signal to noise ratio. They stated that an ensemble of classifiers performed better under these constraints.

In recent years, the research interests shifted towards the use of compressed sensing [6] and compressed learning [1, 3] to study fMRI data in neuroscience applications. Since compressed sensing (CS) provides sparsity sampling approach that allows samples of small number of nonzero coefficients represented within data (or image) be used to reconstruct the original data, it can provide computationally efficient techniques. In other words, it can help the detection of nonzero coefficients that hold most of the important information of the data and the reconstruction of the data (signal or image) with a fewer samples than required by Shannon-Nyquist sampling theorem [5] for privacy protection purposes.

To describe compressed sensing mathematically, let's assume we have a vector $\mathbf{x}$, which represents the data (signal or an image)

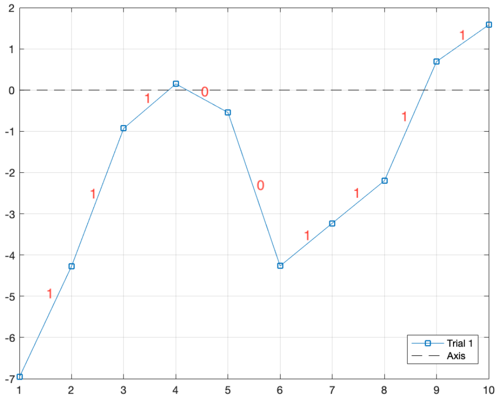

Figure 1: Markov Chain Example: It illustrates a segment of a two-state Markov Chain and its state changing characteristics using the data associated with the Star Plus dataset.

that has N samples or pixels. Now suppose y represents the K ($< N$) measurements that are allowed to be used to optimize an objective function (e.g., an information gain or an entropy), then to obtain y, a matrix $\mathbf{A}$ is needed, which satisfies the mathematical relationship, $y = \mathbf{Ax}$, where the matrix $\mathbf{A}$ is called the sensing matrix and it allows the obtainment of y from $\mathbf{x}$. So $\mathbf{x}$ is $N \times 1$ vector that represents the original signal or image, y is a $K \times 1$ vector of the output, and $\mathbf{A}$ is the $K \times N$ sensing matrix, or how the output is acquired [6].

Similarly, compressed learning (CL) occurs when predictive models are applied to the compressed data to see if accuracy scores are at least similar if not better than when used on the original data. Another way to define CL is as a mathematical framework that incorporates CS with machine learning models. Compared to CS, compressed learning is focused on the interference of the signal rather than the reconstruction of the signal as in CS. Some of the applications of CL is in compressive image classification [23], compressive acquisition of dynamic scenes [17], compressive watermark detection [20], and compressive hyper-spectral image analysis [8].

Recently, in 2016, Adler et al. [1] were able to show for the first time the use of deep neural network models on compressive sensing and non-linear interference. The neural network was also able to optimize the compressed sensing matrix and inference operator, which led to significant improvements in image classification. This process was done through the use of deep learning models that optimize the sensing matrix and the inference operator. Their approach had the first layer of the convolution neural network (CNN) perform the sensing matrix part, thereafter the following layers would perform the inference stage. After the network has been trained, the first layer can be detached from the inference layers. This creates two independent elements in the CL system. After training the network on images of MNIST dataset, the classification error performance was used for sensing rates.

Adler et al. [1] also showed that when compared with Smashed Filter and Random Sensing with CNN models, their deep learning model outperformed both, especially as the sensing rate and number of measurements increased. They also claimed the improvement is due to the optimization of the sensing matrix in the proposed deep learning approach over other approaches that use the standard sensing matrix. Since the use of compressed sensing and compressed learning showed successful results in other applications, the recent research in neuroscience has adopted these techniques. Focusing on the application to fMRI data, Carmi et. al. [4] proposed a isometric transformation that helps the data to follow those properties and a Bayesian-based compressed sensing to extract complete statistical information for the estimated parameters. These approaches helped them achieve higher fMRI classification accuracy. In 2011, Lee et. al. [10] used compressed sensing to recover sparse brain network using fMRI data. They estimated the partial correlation, as a compressed sensing approach, using the penalized linear regression and used that estimation to optimize the recovery of the sparse brain connectivity. They controlled the sparsity through sensing the correlation matrix using a threshold.

Similarly, Yan et. al. [21], in 2013, focused on optimizing the quality of fMRI image, when reconstructed under the measurement constraints. They considered fMRI data as a sequence that is generated by a linear dynamical system. They assumed the fMRI images are sparse over time (i.e. compressed sensing) in the wavelet domain, and then utilized the correlation properties of the adjacent fMRI images to extract measurements that are optimal for the fMRI image reconstruction. However, predominantly, the current research efforts are towards developing techniques, including machine learning approaches, that are efficient in extracting mental or cognitive states of brain using fMRI data, or reconstructing the images. No significant research work can be seen in neuroscience research literature for developing privacy-preserving predictive models and algorithms. Nevertheless, to the best of our knowledge, the research focusing on data privacy in fMRI data analytics for neuroscience applications, using compressed sensing and compressed learning with two-state Markov chain is yet to be explored.

In a recent research, we have performed a preliminary study on compressed sensing [6] and compressed learning [1, 3] with two-state Markov chain to characterize the transition behavior of fMRI signals. We used these transition states to construct compressed sensing matrix and transformed the signals for compressed learning to build a privacy-preserving predictive model. This approach served as a feature selection mechanism. We have presented this work with our preliminary results and findings at the Stanford Compression workshop (DOI: 10.13140/RG.2.2.16571.87849). This model showed its strengths with Logistic Regression (LR) and Sparse Multinomial Logistic Regression (SMLR), while showing significant weaknesses with Naïve Bayes (NB) and Artificial Neural Network (ANN). However, we observed that our approach, as a feature selection mechanism, eliminates the distinguishable characteristics between the classes, which led to drawbacks for the techniques (NB and ANN) that rely heavily on statistical nature of the data.

2 THE CORE IDEA

The core idea that is executed in the technique presented in this paper is the adoption of the power of compressed sensing and compressed learning techniques to improve privacy protection of the

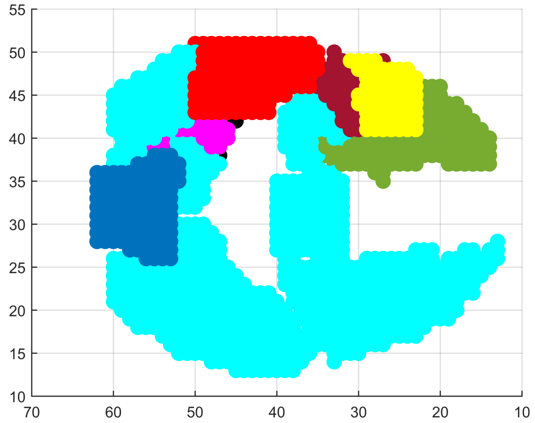

Figure 2: Sufficient ROIs: It shows the 7 ROIs recommended by Star Plus project for the subject numbered 04847.

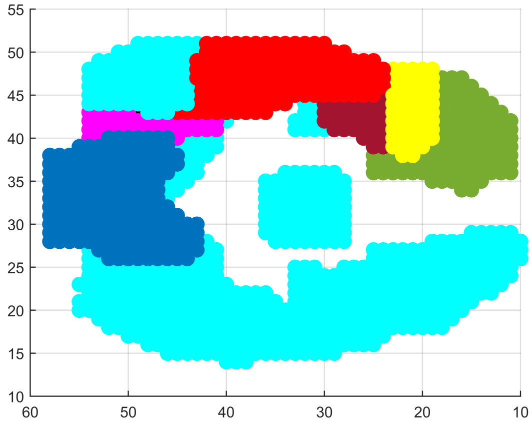

Figure 3: Sufficient ROIs: It shows the 7 ROIs recommended by Star Plus project for the subject numbered 05675.

fMRI predictive models through the introduction of compressed constraints in the underdetermined linear systems. These compressed constraints are generated using transition probabilities of a signal-dependent or signal-independent Markov Chain for the compressed sensing matrix. The fMRI privacy, considered in this work, is defined by the uniqueness of brain's subregions (a subset of region-of-interests) that contribute to the development of a brain network a biometric, when an individual reacting to a stimulus.

The unauthorized discovery of such subregions is considered as privacy vulnerability, and that is the center of attention of this paper. Our goal is to use Markov Chain to construct this compressed sensing matrix such a way that the inherent sparsity is altered to induce confusion and diffusion (the two components of the traditional cryptographic techniques) for privacy protection.

3 PROPOSED COMPUTATIONAL APPROACH

A true predictive model for fMRI data analytics can be mathematically defined as a simple parametric model as follows:

$$y = f_\beta(\mathbf{x}), \tag{1}$$

where y is the response (prediction) variable, f is the predictive model, β is the model parameters, and the vector $\mathbf{x}$ is the set of predictors which form a feature space. The goal of a predictive modeling is to find the best combination $(\hat{f}, \hat{\beta})$ of the model f and the parameter β for a given data set of $(y, \mathbf{x})$ such that the error $\epsilon_1(> 0)$ measured by the quality metric $\rho(y, \hat{y})$ is minimum, where $\hat{y}$ is the predicted respondes by the derived model $\hat{f}$. Such a learning model, if it is highly efficient, can easily disclose the connection between the features. This type of revelation of connections in fMRI data helps the extraction of the brain networks that are responsible for the individual's thoughts and opinions.

In our proposed computational approach, we define a parametrized compressed sensing and compressed learning model as follows:

$$y_{\mathbf{A}} = f_\beta(\mathbf{x'}), \tag{2}$$

where $\mathbf{x'} = \mathbf{Ax}$ and $\mathbf{A}$ is the compressed sensing matrix (we call CS-matrix) that is computed using a two-state Markov chain modeling in our approach. The goal of our proposed predictive modeling is to find the best model $\hat{f}$, the parameter $\hat{\beta}$ and the matrix $\hat{A}$ combination $(\hat{f}, \beta, \hat{A})$ for a given set $(y, \mathbf{x})$ such that the error $\epsilon_2(> 0)$ measured by the quality metric $\rho(y, \hat{y}_A)$ is minimum and $|\epsilon_1 - \epsilon_2| < \epsilon$, where $\epsilon > 0$. Also, note that the purpose of the matrix $\mathbf{A}$ is not dimensionality reduction, rather it uses all the features but creates compressed sensing environment, which will allow the induction of confusion on the communicating ROI through compression. As mentioned earlier, compressed sensing is a sparsity sampling method which makes the original data $\mathbf{x}$ compressed and used in compressed learning for building predictive models. The novelty of our proposed approach is through the use of a two-state Markov chain [15] and its transition probabilities for compressed sensing and learning. We are venturing to ensure that the trade-off between the prediction accuracy and data privacy are optimized after compressed sensing has been applied.

3.1 Computing CS-matrix

The computation of CS-matrix is performed for every observation (i.e., each trial) that reflects fMRI signal that is associated with an individual (or a human subject). Suppose the observation $\mathbf{X}$ is represented by the vector $\mathbf{X} = (\mathbf{x}, x_{q+1})$, where the vector $\mathbf{x} = (x_1, x_2, ...x_q)$ and $\{x_i, i = 1, 2, \ldots, q+1\}$ is a stochastic process, then we treat it as a Lévy process by defining the differences $\{x_{i+1} - x_i, i = 1, 2, \ldots, q\}$ are independent and they follow the same distribution that depends only on the time difference. Then we define a sequence a_i of 0s and 1s as follows:

$$a_i = \begin{cases} 0 & x_{i+1} - x_i < 0 \\ 1 & x_{i+1} - x_i \geq 0 \end{cases}$$

Hence the sequence $\{a_i, i = 1, 2, \ldots, q\}$ forms a Markov chain. An example of a segment of such a process that is generated by using the CMU data set is presented in Figure 1. This sequence

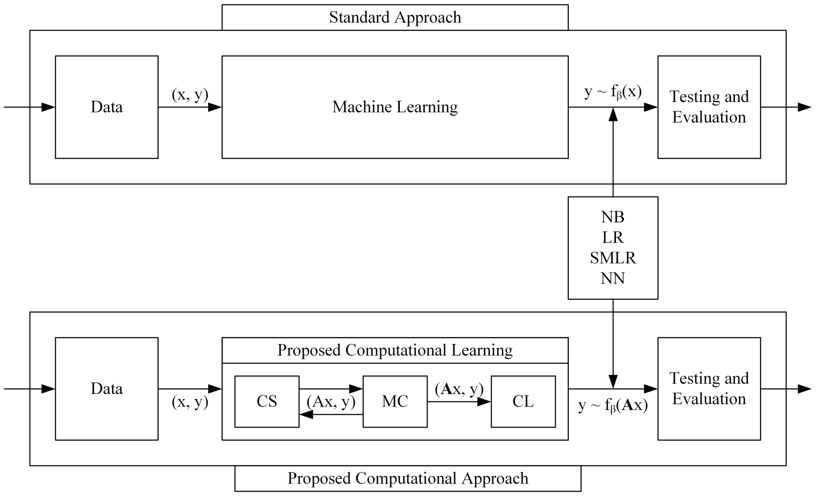

Figure 4: Experimental framework: It shows two experimental processes. The first process (top-half) is the process of evaluating the predictive models (NB, LR, SMLR, and NN) using the standard machine learning. The second process (bottom-half) is the process of evaluating the predictive models (NB, LR, SMLR, and NN) using the proposed computational approach.

(Markov chain) of 0s and 1s is used to construct the CS-matrix $\mathbf{A}_{1\times q}$ for our proposed computational technique, where the i^{th} element of the CS-matrix $\mathbf{A}$ is a_i. Therefore, the compressed form **x'** of the fMRI signal **x** can be derived as follows:

$$\mathbf{x'} = \mathbf{A} * \mathbf{x}, \tag{3}$$

where the operator $*$ defines the element-by-element matrix multiplication; hence, the dimensions of **x** and **x'** are the same. In our proposed work, we have explored two types of compressed sensing, the first one what we call is signal-independent compressed sensing and the second one is signal-dependent compressed sensing. We also explored another approach what we call is sequence cloning. They are described below.

3.1.1 Sensing Types. For the first type, we use the first observation from the data for computing the CS-matrix, and then apply it to all the observations to obtain the compressed observations. For the second type, we compute the CS-matrix for an observation based on its own CS-matrix that is computed using its stochastic behavior, and then use it to generate compressed observation. Suppose, there are n observations, $\mathbf{x}_j$ in the dataset, then we construct n CS-matrices $\mathbf{A}_j, j = 1, 2, \ldots, n$, where the j^{th} CS-matrix $\mathbf{A}_j$ is constructed using the observation $\mathbf{x}_j$, and the compressed observation of $\mathbf{x}_j$ is obtained as follows:

$$\mathbf{x'}_j = \mathbf{A}_j * \mathbf{x}_j, \tag{4}$$

3.1.2 Sequence Cloning. As a result of the Markov chain, for each observation (or a trial), we can create separate transition probability matrix P with the probabilities of a p_{00}, p_{01}, p_{10}, and p_{11}:

$$A = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11}, \end{bmatrix} \tag{5}$$

where the probability $p_{ij}, i, j \in 0, 1$ indicates the transition probability of the sequence $\{a_i, i = 1, 2, \ldots, q\}$ from the state i to the state j. These transition probabilities facilitate the generation of multiple sequences for the CS-matrix associated with an observation. We call this process a sequence cloning which in turn helps us generate multiple compressed observations, say k, for every observation.

3.2 StarPlus fMRI Dataset

In this research, we have utilized the fMRI datasets that are publicly available at the Carnegie Mellon University's StarPlus fMRI data site http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/. Some background on the data and the study are discussed in this section. One of the goals of CCBI, as stated at http://www.ccbi.cmu.edu/, was to explain how thought emerges from brain function and if brain dysfunctions show any affect on thoughts. When looking at the data there were systematic steps that were taken with respect to data setup. To allow for the evaluation of balanced and unbalanced data sets, both the conditions labeling and first stimulus labeling sets were included for this experiment. This allowed for deep evaluation of the models to see if there were any changes in the classification accuracy when using balanced and unbalanced data sets. We have divided their description into three categories (Data Acquisition Descriptors, Data File Descriptors, and Stimulus Label Descriptors) for the convenience for our analysis and discussed subsequently.

3.2.1 Data Acquisition Descriptors. The StarPlus experiment consisted of a set of trials, which created data that was also partitioned into trials. For select intervals, the participant was at rest or gazed at a fixation point on the screen in front of them. For select other trials, the participant was shown a picture and sentence, and then was instructed to press a button if the sentence described the picture. Half of the trials started with the picture shown first and the other half started with the sentence shown first. The timing of the trials was setup as: the first stimulus is shown, this could be the sentence or picture. Following this, four seconds later, the stimulus was removed and replaced by a blank screen. Four seconds

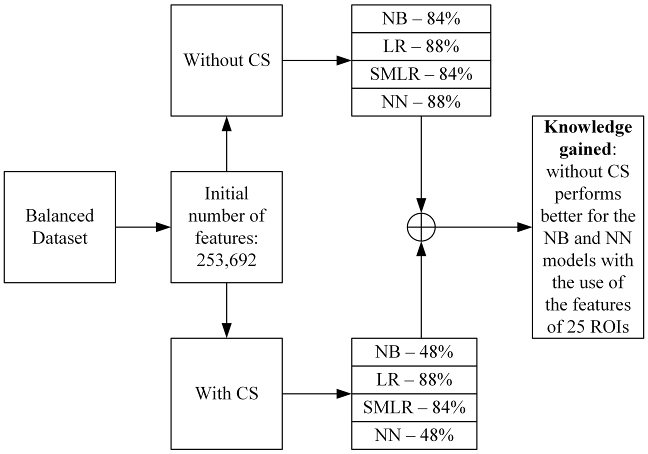
Without CS
NB – 84%
LR – 88%
SMLR – 84%
NN – 88%
Knowledge gained: without CS performs better for the NB and NN models with the use of the features of 25 ROIs
Balanced Dataset
Initial number of features: 253,692
With CS
NB – 48%
LR – 88%
SMLR – 84%
NN – 48%

Figure 5: It presents the stage 1 prediction accuracy results for NB, LR, SMLR, and NN using the benchmark and the proposed approaches, when balanced data is used. It also shows our incremental understanding of the approach.

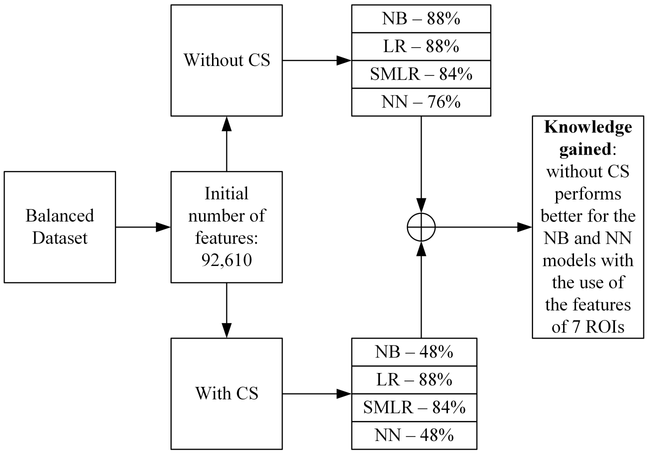
Without CS
NB – 88%
LR – 88%
SMLR – 84%
NN – 76%
Knowledge gained: without CS performs better for the NB and NN models with the use of the features of 7 ROIs
Balanced Dataset
Initial number of features: 92,610
With CS
NB – 48%
LR – 88%
SMLR – 84%
NN – 48%

Figure 6: It presents the stage 2 prediction accuracy results for NB, LR, SMLR, and NN using the benchmark and the proposed approaches, when balanced data is used. It also shows our incremental understanding of the approach.

after that, the second stimulus was presented to the participant and remained on the screen for four seconds or until the participant pressed the button, whichever of the two came first.

Following the removal of the second stimulus from the screen, there was a 15 second rest period to end the trial. All together each of the trials lasted for 27 seconds; 4 sec + 4 sec + 4 sec + 15 sec = 27 seconds. As the experiment progresses, fMRI images were taken and recorded every 500 msec bringing the total images for each trial to 54. This total number of images is calculated by 2 x 27 = 54 images. The data was then marked up into 25 defined region of interests (ROIs), yet it is recommended by StarPlus to use only seven ROIs, when trying to look at classification on the data. These seven ROIs are 'CALC', 'LIPL', 'LT', 'LTRIA', 'LOPER', 'LIPS', and 'LDLPFC' - examples of the seven ROIs of two subjects in the StarPlus fMRI dataset, are shown in Figures 2 and 3. From these figures, we can clearly see the distinction between the shapes, sizes, and locations of the voxels belonging to these seven ROIs for different subjects. The voxels in these regions form the functional brain networks based on the stimulus shown at that time.

Therefore, the goal of our privacy-preserving predictive model is to identify a subset of voxels that help predictive models achieve higher accuracies, while creating confusion and diffusion which will prevent from unauthorized users (or attackers) find approaches to reconstruct the original seven regions with the same shapes, sizes, and locations, so that the subjects identity can be easily extracted.

3.2.2 Data File Descriptors.

The dataset contains six Matlab data files for six different subjects. Each data file has a format of "data-starplus-04847-v7.mat", the format defines "data" as data file, "starplus" as StarPlus experiment, and "04847" signifies the unique number for the subject. This means that this data file only pertains to the subject 04847. Once this file is loaded, there are three variables that are defined: data, info, and meta. The variable data contains the data values for the image intensity values, meta contains the metadata on the data set and the time series for the images, defined above, are appointed as trials, and the info variable describes each of these trials. The variable data consist of the observed data. The data structure 'data' is a 54x1 cell array, where each cell is one trial. Each element in the cell array is $N \times V$ of the fMRI observations. To better understand and define the 'data' structure, the element data{$\mathbf{x}$}(t,v) (what we call in our experiment the "features") returns the fMRI observation of voxel v, at time t within $\mathbf{x}$.

Now when looking at subject 04847, the dimension of the data are 50 x 253,692. Four trials were dropped when initially building the data dimensions, so the trials are now 50. 253,692 comes from the number of images collected during each trial, 54, and the number of voxels for the subject, which for subject 04847 is 4,698. When the 54 images are multiplied by the 4,698 voxels this gives the feature space of 253,692. This 253,692 is captured when using all 25 ROIs, but when the ROIs are reduced to the 7 previously discussed ('CALC', 'LIPL', 'LT', 'LTRIA', 'LOPER', 'LIPS', and 'LDLPFC') the number of voxels is reduced to 1,715. Now, when using the reduced number of voxels, 1,715, multiplied by the number of images in each trial, 54, the feature space is now 92,610.

3.2.3 Stimulus Label Descriptors.

One final point on the data is to discuss the conditions and first stimulus labeling. These labels were utilized as the label sets for the data's training and testing sets. Both of these fields can be found within the info variable. Info is a 1x54 structure array, which holds much of the data that describes the 54 different time intervals (trails). The conditions label has values of 0, 1, 2, and 3. A condition value of 0 means the data in this segment should be ignored. A value of 1 means the segment is during a time when the subject is at rest or it is a fixation interval. Condition value of 2 signifies that it is a sentence or picture trial, where the sentence is not negated. Finally, a condition value of 3 signifies it is a sentence or picture trial where the sentence is negated.

The first stimulus label has values of 'P' or 'S', which indicate whether the 'picture' was displayed first before the 'sentence' (first stimulus = 'P') or whether the 'sentence' was displayed first before

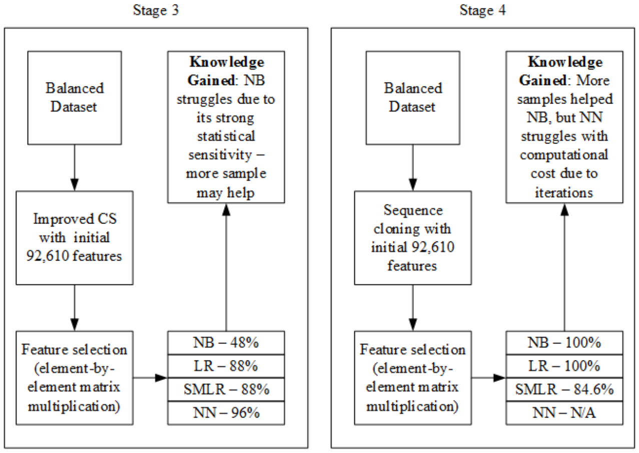
Stage 3
Balanced Dataset
Improved CS with initial 92,610 features
Feature selection (element-by-element matrix multiplication)
NB – 48%
LR – 88%
SMLR – 88%
NN – 96%
Knowledge Gained: NB struggles due to its strong statistical sensitivity – more sample may help
Stage 4
Balanced Dataset
Sequence cloning with initial 92,610 features
Feature selection (element-by-element matrix multiplication)
NB – 100%
LR – 100%
SMLR – 84.6%
NN – N/A
Knowledge Gained: More samples helped NB, but NN struggles with computational cost due to iterations

Figure 7: It presents the stages 3 and 4 prediction accuracy results for NB, LR, SMLR, and NN using the benchmark and the proposed approaches, when balanced data is used. It also shows our incremental understanding of the approach.

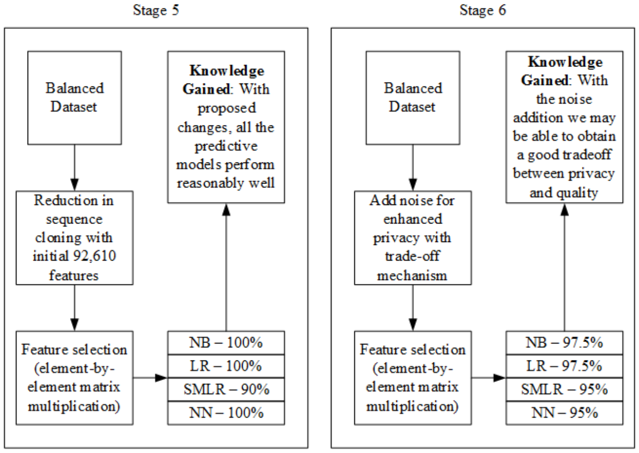
Stage 5
Balanced Dataset
Reduction in sequence cloning with initial 92,610 features
Feature selection (element-by-element matrix multiplication)
NB – 100%
LR – 100%
SMLR – 90%
NN – 100%
Knowledge Gained: With proposed changes, all the predictive models perform reasonably well
Stage 6
Balanced Dataset
Add noise for enhanced privacy with trade-off mechanism
Feature selection (element-by-element matrix multiplication)
NB – 97.5%
LR – 97.5%
SMLR – 95%
NN – 95%
Knowledge Gained: With the noise addition we may be able to obtain a good tradeoff between privacy and quality

Figure 8: It presents the stages 5 and 6 prediction accuracy results for NB, LR, SMLR, and NN using the benchmark and the proposed approaches, when balanced data is used. It also shows our incremental understanding of the approach.

the 'picture' (first stimulus = 'S'). It is also important to note that during this study, to make the first stimulus labeling set similar to the conditions labeling set and to also ensure there were no problems during the classification process, the first stimulus labeling set was changed to 'P' = 1 and 'S' = 2.

4 EXPERIMENTAL ANALYSIS

We systematically performed the experiment to understand the effect of CS-matrix on the compressed sensing and compressed learning tasks towards improving the performance of predictive models; hence, we divided the experiment into six stages. The construction of generalized CS-matrix that is suitable for many predictive models is very challenging. Therefore, the use of our six stages in developing the results and understanding the effect of CS-matrix can help us refining the proposed computational approaches, progressively. An experimental framework is illustrated in Figure 4. The main objective of this framework is the analysis and evaluation of the performance of the predictive models, (Naïve Bayes (NB) [19], Logistic Regression (LR) [13], Spare Multinomial Logistic Regression (SMLR) [9], or 2-layer Feedforward Neural Network (NN) [22]), using the standard machine learning approaches [18] and the proposed computational approaches.

The top-half of the process diagram presented in Figure 4 illustrates this standard process. We also considered one subject only in the analysis presented in this paper. The first step of this process is to prepare the data for the model analysis. This included the reduction of the size of the data table from 54 x 253,692 to 50 x 253,692, when the complete feature space with 25 ROIs is used, and from 54 x 92,610 to 50 x 92,610, when the subspace with 7 ROIs is used. It means the number of trials (or observations) considered for the subject is reduced from 54 to 50. The decrease from 54 to 50 trials is resulted from dropping of any trial with less than 54 images, when we construct our data table. The process presented in the bottom-half of Figure 4 illustrates the use of compressed sensing and compressed learning with the two-state Markov Chain. From here the two-state Markov chain is created using the transition properties of the 50 x 253,692 (when 25 ROIs are used) and 50 x 92,610 (when 7 ROIs are used) data table. This step creates our compressed sensing matrix and then is multiplied by the data matrix (or table). This allows for the spikes in the data to remain, while removing the low points in brain activations. By doing this, we are ensuring that the higher fMRI brain activations are kept and used in the compressed learning models.

The following sections will explain the steps taken in each stage, the reasoning of these steps and any problems faced. Figures 5 to 8 present the 6 stages of the proposed experiment using the balanced data set in section 4.1, and Figures 10 to 13 present the 6 stages of the proposed experiment using the balanced data set in section 4.2.

4.1 Analysis with balanced data

4.1.1 Stage 1: In this stage, we analyzed the standard machine learning and the proposed computational learning (using the signal-independent CS-matrix) approaches using all the 25 ROIs (i.e, with 253,692 features). We first computed the CS-matrix for compressed sensing using the Markov chain sequence generated by the signal-independent strategy, and compressed the data using the matrix. We then applied Naïve Bayes (NB), Logistic Regression (LR), Sparse Multinomial Logistic Regression (SMLR), and neural network (NN) models to the compressed data for compressed learning, and calculated prediction accuracies. We also obtained benchmark results to compare our approach using the four models NB, LR, SMLR, and NN without compressed sensing. The results are presented in Figure 5.

As we can see from the figure that the standard approach (i.e., without compressed sensing) resulted in accuracies 84% (NB), 88% (LR), 84% (SMLR) and 88% (NN), and when compressed sensing was applied, the accuracies were 48% (NB), 88% (LR), 84% (SMLR), and 48% (NN). The results indicate that both approaches perform

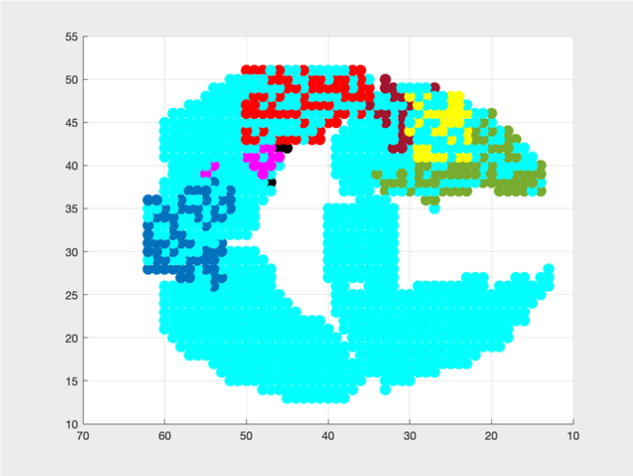

Figure 9: A subset of 7 ROIs: Sensed and compressed subspace for higher prediction accuracy and privacy protection.

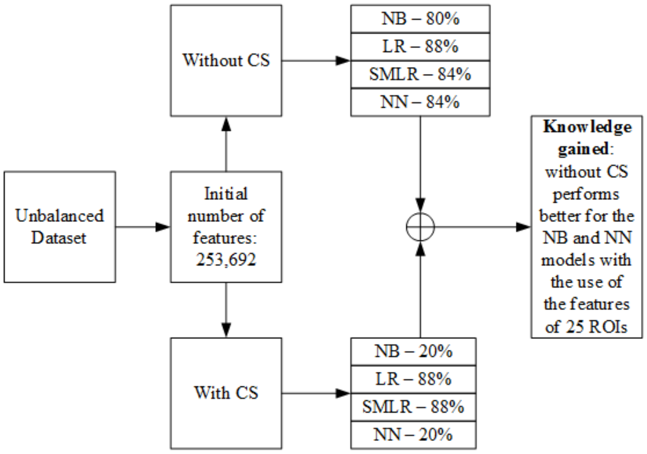
Without CS
NB – 80%
LR – 88%
SMLR – 84%
NN – 84%
Unbalanced Dataset
Initial number of features: 253,692
Knowledge gained: without CS performs better for the NB and NN models with the use of the features of 25 ROIs
With CS
NB – 20%
LR – 88%
SMLR – 88%
NN – 20%

Figure 10: It presents the stage 1 prediction accuracy results for NB, LR, SMLR, and NN using the benchmark and the proposed approaches, when unbalanced data is used. It also shows our incremental understanding of the approach.

equally well with the predictive model LR and SMLR, but the proposed approach performs poorly with the models NB and NN. What we learned from this experiment is that there were many irrelevant features among 253,692 (25 ROIs) that negatively affects the performance of the predictive models. Therefore, we decided to adopt the CMU's StarPlus advice and investigated by limiting the number of ROIs to the suggested seven ROIs, which provided 92,610 features for the rest of our experimental analysis.

4.1.2 Stage 2: In this stage, we also analyzed the standard machine learning and the proposed computational learning (using the signal-independent CS-matrix) approaches, but used 7 ROIs (i.e, with 92,610 features), suggested by the StarPlus fMRI website. We now computed the CS-matrix for compressed sensing using the Markov chain sequence generated by the signal-independent strategy (as before), and compressed the data using the new CS-matrix. We then applied Naïve Bayes (NB), Logistic Regression (LR), Sparse Multinomial Logistic Regression (SMLR), and neural network (NN) models to the compressed data for compressed learning, and calculated prediction accuracies. We also obtained a benchmark results, as previously obtained, to compare our approach using the four models NB, LR, SMLR, and NN without compressed sensing. The results are presented in Figure 6.

It is interesting to see from the figure that the standard approach (i.e., without compressed sensing) resulted in accuracies, 88% (NB), 88% (LR), 84% (SMLR) and 76% (NN) that indicate a slightly increased accuracy for NB, and a significant decrease for NN. However, when the proposed approach with compressed sensing and learning was applied, the accuracies were remained at the same level of accuracies, 48% (NB), 88% (LR), 84% (SMLR), and 48% (NN). Since the results didn't show significant discrepancies overall, we decided to adopt the 7 ROIs for the rest of the experimental analysis, in which several enhancements were integrated to the proposed computational approach. From this experiment, we learned that the use of signal-independent strategy propagates the statistical characteristics of the first signal through other signals and alters their own statistical characteristics; therefore, we used the signal-dependent strategy in the subsequent experimental analysis.

4.1.3 Stage 3: In this stage, to combat the aforementioned problem, we generated individualized Markov Chain sequences for each of the 50 observations (as signal-dependent strategy) and constructed the CS-matrix. Then used this matrix for compressed sensing and analyzed the standard machine learning and the proposed computational learning (using the signal-dependent CS-matrix) approaches, along with the use of 7 ROIs (i.e, with 92, 610 features). The results of this experiment are presented in the left diagram in Figure 7.

Since we have individualized CS-matrix elements (rows of elements) for compressed sensing, we used them to compress the data observation-by-observation separately. We then applied Naïve Bayes (NB), Logistic Regression (LR), Sparse Multinomial Logistic Regression (SMLR), and neural network (NN) models to the compressed data for compressed learning, and calculated prediction accuracies. For the purpose of benchmark results, we selected the best accuracies for the standard approach by combining the results obtained in stage 1 and stage 2 for this approach. Hence, we used 88% (NB), 88% (LR), 84% (SMLR), and 88% (NN) for the standard approach. When the proposed approach with compressed sensing and learning was applied, we were able to obtain the accuracies 48% (NB), 88% (LR), 88% (SMLR), and 96% (NN). These results (in the left diagram of Figure 7) indicate that the proposed approach and the standard approaches perform equally good for the LR and SMLR predictive models, but the proposed approach performed very good for NN, while performing poorly for NB.

What we learned from this experiment is that the Naive Bayes performed poorly, because it is statistically sensitive to variability because of its assumption that the features are independent and the environment is probabilistic. Therefore, it requires multiple samples to estimate the statistical parameters more accurately. As such, we used the sequence cloning technique to generate multiple Markov

Chain sequences for each observation to construct CS-matrix. This is our goal in stage 4.

4.1.4 Stage 4: The transition properties were investigated by calculating the transition probabilities (as per Equation 5) and then from there calculating the transition probabilities for each observation. Now having a transition probability matrix for each observation, we began by cloning 50 new sequences for each observation, creating a new sensing matrix with 2,500 X 92,610 dimensions. Having a new sensing matrix approach, we again tested using the four models with compressed learning and obtained great improvements in each predictive model, 100% (NB), 100% (LR), 84% (SMLR). Yet when using the neural network model, this new sequence cloning approach seemed to be quite resource intensive; hence, simulation was not completed within the acceptable duration. Therefore, we marked the result as “N/A” and started to seek for an improvement. What we have learned from this experiment is that NN may not require all the 50 sequences to adopt the compressed sensing due to its stochastic gradient descent optimization; hence, we looked at the option of reducing the number of cloned sequences empirically. This is the next stage of our experimental analysis.

4.1.5 Stage 5: Reduced Sequence Cloning. The number of sequence clones were decreased to determine how low the number of clones could be while still maintaining a high accuracy for all predictive models. The optimal number proved to be 4 sequence clones, empirically; this allowed for the accuracies to remain eminently high 100% (NB), 100% (LR), 90% (SMLR), and 100% (NN)) and reduced the problem of resource lagging, when using the neural network model. These results are presented in the left diagram of Figure 8. The significant improvements in the prediction accuracies leaves us enough room to increase the privacy protection further through magnitude degradation and compressed sensing which will induce the confusion and diffusion to the unauthorized (or an attacker) individual in determining which voxels are active and communicating in the brain networks.

At this stage, to see the effect of compressed sensing, we have highlighted the sub ROIs that are identified by the compressed sensing matrix. Figure 9 displays these regions. These are the subregions, within each ROI, that are used to improve the above prediction accuracies for the four models. These regions cannot be obtained without the CS-matrix that was used; hence, it is difficult to reconstruct the original functional brain network by the unauthorized personnel. Since we obtained very high prediction accuracies, now we were able to degrade the edge magnitudes of the brain networks, which will allow possible confusion in determining which voxels are communicating. This is the goal of stage 6.

4.1.6 Stage 6: Trade-offs with data degradation. Now that performance was no longer an issue and the accuracy results were profoundly satisfactory, noise was added to the original data to generate a trade-off mechanism between classification accuracy and privacy protection. The degradation of the data added additional privacy preserving measures, while still maintaining high classification accuracy. The results of this stage was 97.5 (NB), 97.5 (LR), 95 (SMLR), and 95 (NN). Now this proposed model was able to balance and address privacy concerns within the fMRI data while maintaining high classification accuracy results showing promising future utilization. Now that performance was no longer an issue and the accuracy results were profoundly satisfactory, noise was added to the original data to generate a trade-off mechanism between classification accuracy and privacy protection. The degradation of the data added additional privacy preserving measures, while still maintaining high classification accuracy. The results of this stage was 97.5 (NB), 97.5 (LR), 95 (SMLR), and 95 (NN). Now this proposed model was able to balance and address privacy concerns within the fMRI data while maintaining high classification accuracy results showing promising future utilization.

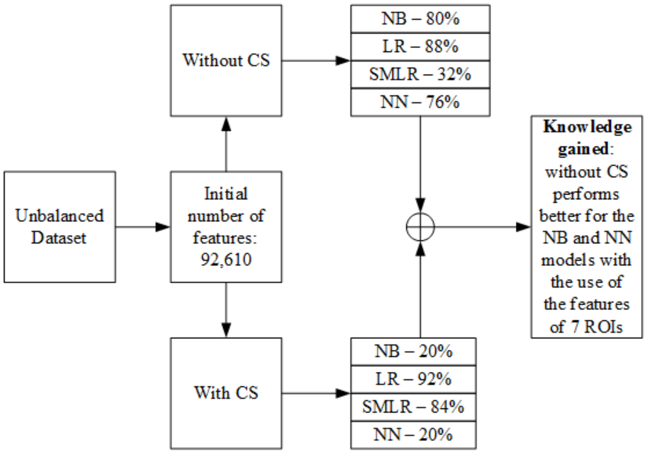
Without CS
NB – 80%
LR – 88%
SMLR – 32%
NN – 76%
Knowledge gained: without CS performs better for the NB and NN models with the use of the features of 7 ROIs
Unbalanced Dataset
Initial number of features: 92,610
With CS
NB – 20%
LR – 92%
SMLR – 84%
NN – 20%

Figure 11: It presents the stage 2 prediction accuracy results for NB, LR, SMLR, and NN using the benchmark and the proposed approaches, when unbalanced data is used. It also shows our incremental understanding of the approach.

4.2 Analysis with unbalanced data

The same steps of the stages 1 to 6 in the experiment with the balanced dataset were carried out now using the unbalanced dataset in this experiment. The corresponding results are presented in Figures 10 to 13, respectively. As we can see from these figures, very similar results were obtained; however, the effect of the unbalanced nature of the data can still be seen from the small difference displayed in the prediction accuracies.

Therefore, based on all of these results, we can select the proposed approach with the reduced cloning for all the four predictive models and the unbalanced data with the accuracies around 100% for NB, LR, and NN, and 90% for SMLR. Similarly, if the degradation approach is adopted for trade-off between quality and privacy, then we have accuracies 92.5% for NB, 87.5% for LR, 82.5% for SMLR, and 97.5% for NN. As an end result, we can see NN performs better overall in both balanced and unbalanced scenarios, and trade-off or “no” trade-off options.

4.3 Discussion on Privacy Strength

In this paper, we have not looked into any numerical measures to quantify the fMRI data privacy, which is one of our future goals. Hence, we provide a discussion based on the visual comparison to

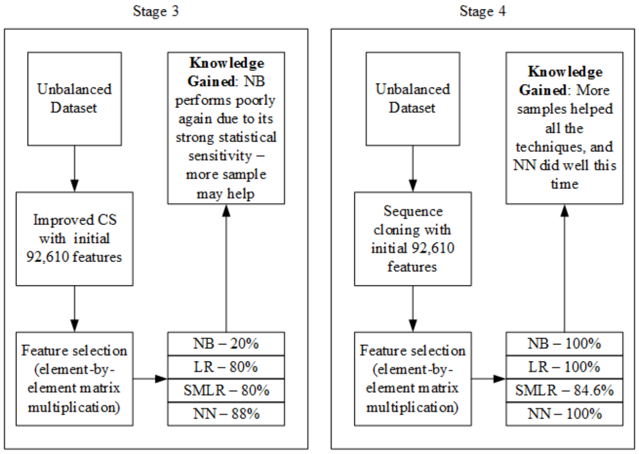
Stage 3
Unbalanced Dataset
Knowledge Gained: NB performs poorly again due to its strong statistical sensitivity – more sample may help
Improved CS with initial 92,610 features
Feature selection (element-by-element matrix multiplication)
NB – 20%
LR – 80%
SMLR – 80%
NN – 88%
Stage 4
Unbalanced Dataset
Knowledge Gained: More samples helped all the techniques, and NN did well this time
Sequence cloning with initial 92,610 features
Feature selection (element-by-element matrix multiplication)
NB – 100%
LR – 100%
SMLR – 84.6%
NN – 100%

Figure 12: It presents the stages 3 and 4 prediction accuracy results for NB, LR, SMLR, and NN using the benchmark and the proposed approaches, when unbalanced data is used. It also shows our incremental understanding of the approach.

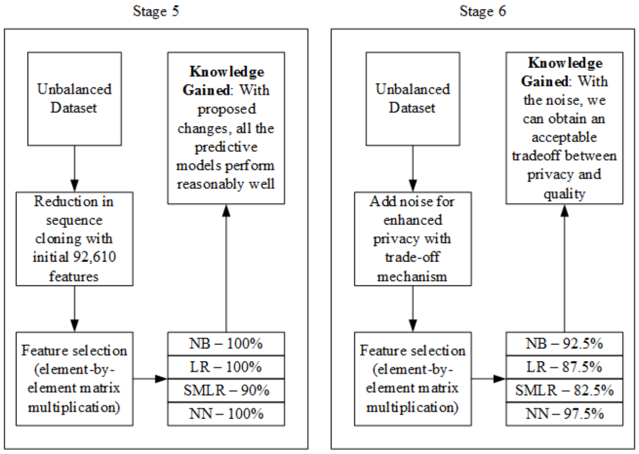
Stage 5
Unbalanced Dataset
Knowledge Gained: With proposed changes, all the predictive models perform reasonably well
Reduction in sequence cloning with initial 92,610 features
Feature selection (element-by-element matrix multiplication)
NB – 100%
LR – 100%
SMLR – 90%
NN – 100%
Stage 6
Unbalanced Dataset
Knowledge Gained: With the noise, we can obtain an acceptable tradeoff between privacy and quality
Add noise for enhanced privacy with trade-off mechanism
Feature selection (element-by-element matrix multiplication)
NB – 92.5%
LR – 87.5%
SMLR – 82.5%
NN – 97.5%

Figure 13: It presents the stages 5 and 6 prediction accuracy results for NB, LR, SMLR, and NN using the benchmark and the proposed approaches, when unbalanced data is used. It also shows our incremental understanding of the approach.

justify the privacy protection capability of the proposed approach. As stated earlier, our goal was to identify a subset of voxels that help predictive models achieve higher accuracies, while creating confusion and diffusion to unauthorized users (or attackers) which will resist them finding approaches that can help them reconstruct the original seven regions with the same shapes, sizes, and locations, so that the subjects identity can be easily extracted.

If we compare, the results in Figures 2, 3, and 9, with the assumption, that the first two images (Figures 2 and 3) are not available to the users, and the subsets that are highlighted in Figure 9 are the only ones available for them to build and evaluate their predictive models, then it is difficult for them to reconstruct the original regions and reveal the identity of the subjects. Note that: since the highlighted subset of voxels are sufficient to test predictive models and achieve higher accuracies, only this subset will be distributed and scattered over multiple platforms to help distributed machine learning research. Hence, the privacy is protected, while accomplishing the tasks of developing and evaluating machine learning models and algorithms.

5 CONCLUSION

The proposed compressed sensing and compressed learning approach, while helping all the four evaluated predictive models improve their prediction accuracies, helps artificial neural network to improve its performance significantly. The proposed computational approach is also capable of identifying subsets of voxels that support the enhancement of privacy protection of predictive models through the integration of confusion and diffusion – the two important properties of any cryptographic algorithms – that lead to the resistance of recovering the original and individualized shapes, sizes, and locations of voxels for the subjects.

In this research, we have not studied the privacy strengths by quantifying through the use of numerical measures; hence, one of our future goals is to explore such measures in this study. We have also not incorporated all the subjects provided by the Star Plus datasets in this study; therefore, we will continue this research with more subjects. We expect some challenges due to data heterogeneity that will be presented in multiple subjects; therefore, we will study this problem using asymptotic transition probabilities of the Markov Chain that we proposed. Finally, based on this study, our conclusion is that there exits a subsets of voxels within seven ROIs that can be used to improve prediction accuracy of a predictive models while making them stronger in protecting the data privacy.

REFERENCES

[1] Amir Adler, Michael Elad, and Michael Zibulevsky. 2016. Compressed Learning: A Deep Neural Network Approach. *arXiv:1610.09615 [cs]* (Oct. 2016). http://arxiv.org/abs/1610.09615 arXiv: 1610.09615.

[2] Carlos Cabral, Margarida Silveira, and Patricia Figueiredo. 2012. Decoding visual brain states from fMRI using an ensemble of classifiers. *Pattern Recognition* 45, 6 (2012), 2064–2074.

[3] Robert Calderbank, Sina Jafarpour, and Robert Schapire. 2009. Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain. *preprint* (2009).

[4] Avishy Carmi, Tara N Sainath, Pini Gurfil, Dimitri Kanevsky, David Nahamoo, and Bhuvana Ramabhadran. 2010. The use of isometric transformations and Bayesian estimation in compressive sensing for fMRI classification. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 493–496.

[5] Gilles Chardon, A Leblanc, and Laurent Daudet. 2011. Plate impulse response spatial interpolation with sub-Nyquist sampling. *Journal of sound and vibration* 330, 23 (2011), 5678–5689.

[6] D.L. Donoho. 2006. Compressed sensing. *IEEE Transactions on Information Theory* 52, 4 (April 2006), 1289–1306. https://doi.org/10.1109/TIT.2006.871582

[7] Nico UF Dosenbach, Binyam Nardos, Alexander L Cohen, Damien A Fair, Jonathan D Power, Jessica A Church, Steven M Nelson, Gagan S Wig, Alecia C Vogel, Christina N Lessov-Schlaggar, et al. 2010. Prediction of individual brain maturity using fMRI. *Science* 329, 5997 (2010), 1358–1361.

[8] Jürgen Hahn, Simon Rosenkranz, and Abdelhak M Zoubir. 2014. Adaptive compressed classification for hyperspectral imagery. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1020–1024.

[9] B. Krishnapuram, L. Carin, M.A.T. Figueiredo, and A.J. Hartemink. 2005. Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 6 (June 2005), 957–968. https://doi.org/10.1109/TPAMI.2005.127

[10] Hyekyoung Lee, Dong Soo Lee, Hyejin Kang, Boong-Nyun Kim, and Moo K Chung. 2011. Sparse brain network recovery under compressed sensing. *IEEE Transactions on Medical Imaging* 30, 5 (2011), 1154–1165.

[11] Tom M Mitchell, Rebecca Hutchinson, Radu S Niculescu, Francisco Pereira, Xuerui Wang, Marcel Just, and Sharlene Newman. 2004. Learning to decode cognitive states from brain images. *Machine learning* 57, 1-2 (2004), 145–175.

[12] Thomas Naselaris, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant. 2011. Encoding and decoding in fMRI. *Neuroimage* 56, 2 (2011), 400–410.

[13] Andrew Y Ng and Michael I Jordan. 2002. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in neural information processing systems*. 841–848.

[14] Francisco Pereira, Tom Mitchell, and Matthew Botvinick. 2009. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45, 1 (2009), S199–S209.

[15] Mark A. Pinsky and Samuel Karlin. 2010. *An Introduction to Stochastic Modeling* (4th edition ed.). Academic Press.

[16] Srikanth Ryali, Kaustubh Supekar, Daniel A Abrams, and Vinod Menon. 2010. Sparse logistic regression for whole-brain classification of fMRI data. *NeuroImage* 51, 2 (2010), 752–764.

[17] Aswin C Sankaranarayanan, Pavan K Turaga, Richard G Baraniuk, and Rama Chellappa. 2010. Compressive acquisition of dynamic scenes. In *European Conference on Computer Vision*. Springer, 129–142.

[18] Shan Suthaharan. 2016. *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*. Springer US. https://www.springer.com/us/book/9781489976406

[19] Jaideep Vaidya, Murat KantarcÄśoÄ§lu, and Chris Clifton. 2008. Privacy-preserving NaÃŕVe Bayes Classification. *The VLDB Journal* 17, 4 (July 2008), 879–898. https://doi.org/10.1007/s00778-006-0041-y

[20] Qia Wang, Wenjun Zeng, and Jun Tian. 2014. A compressive sensing based secure watermark detection and privacy preserving storage framework. *IEEE transactions on image processing* 23, 3 (2014), 1317–1328.

[21] Shulin Yan, Lei Nie, Chao Wu, and Yike Guo. 2013. An approximation approach to measurement design in the reconstruction of functional MRI sequences. In *International Conference on Brain and Health Informatics*. Springer, 115–125.

[22] Naimin Zhang, Wei Wu, and Gaofeng Zheng. 2006. Convergence of Gradient Method with Momentum for two-Layer Feedforward Neural Networks. *Trans. Neur. Netw.* 17, 2 (March 2006), 522–525. https://doi.org/10.1109/TNN.2005.863460

[23] E Zisselman, A Adler, and M Elad. 2018. Compressed Learning for Image Classification: A Deep Neural Network Approach. *Processing, Analyzing and Learning of Images, Shapes, and Forms* 19 (2018), 3–17.