Region-specific fMRI Dictionary for Decoding Face Verification in Humans

Daksha Yadav[1], Naman Kohli[1], Shruti Nagpal[2], Maneet Singh[2], Prateekshit Pandey[3], Mayank Vatsa[2], Richa Singh[2], Afzel Noore[1], Gokulraj Prabhakaran[4], and Harsh Mahajan[5]

[1]West Virginia University, USA, [2]IIIT-Delhi, India, [3]University of Pennsylvania, USA
[4]Otto-von-Guericke University, Germany, [5]Mahajan Imaging, India

***Abstract*—This paper focuses on decoding the process of face verification in the human brain using fMRI responses. 2400 fMRI responses are collected from different participants while they perform face verification on genuine and imposter stimuli face pairs. The first part of the paper analyzes the responses covering both cognitive and fMRI neuro-imaging results. With an average verification accuracy of 64.79% by human participants, the results of the cognitive analysis depict that the performance of female participants is significantly higher than the male participants with respect to imposter pairs. The results of the neuro-imaging analysis identifies regions of the brain such as the left fusiform gyrus, caudate nucleus, and superior frontal gyrus that are activated when participants perform face verification tasks. The second part of the paper proposes a novel two-level fMRI dictionary learning approach to predict if the stimuli observed is genuine or imposter using the brain activation data for selected regions. A comparative analysis with existing machine learning techniques illustrates that the proposed approach yields at least 4.5% higher classification accuracy than other algorithms. It is envisioned that the result of this study is the first step in designing brain-inspired automatic face verification algorithms.**

I. Introduction

The processing capabilities of the human brain have fascinated the researchers for the past few decades, leading to attempts of better understanding and emulating the brain's functionality. Specifically, the face recognition capabilities of humans has motivated researchers for developing *intelligent* algorithms for automated matching. Several algorithms, ranging from Gabor transform to deep learning based architectures, have been proposed, that seek to emulate the complex working of the human brain [1]–[6].

Parallely, in cognitive neuroscience, researchers have followed independent research directions to understand brain functioning by means of behavioral (cognitive) and sense based approaches such as electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI) [7]. In experiments pertaining to face recognition, researchers have focused on understanding *face perception* and brain areas which are responsible for such tasks. For instance, researchers have tried to unravel the face discrimination abilities present in newborns and monkeys [8]. It has also been established that fusiform gyrus and lingual gyrus are responsible for face perception [7], [9], [10]. Similarly, cognitive neuroscientists have demonstrated that humans have innate capabilities for recognizing familiar faces even in the presence of moderate disguise, makeup, and occlusion [11]–[13]. These results have aided the biometrics (face) research community to develop novel algorithms for challenging tasks such as plastic surgery [14], [15] and unconstrained face recognition [5].

In order to move closer to building brain-inspired algorithms for face recognition, it is required to understand various facets of brain processing. There are several questions that require exploration; questions related to the processing capabilities of the human brain and the task of face recognition. For instance, (i) what regions are involved for face perception, (ii) what ancillary information is required for the task of face recognition, (iii) what is the effect of familiarity, memory, gender, and race for the task of face recognition, and (iv) how the brain processes *genuine* and *imposter* face pairs[1]. Traditionally, experimental design in cognitive neuroscience studies focus on self face recognition, familiar face recognition, or kin face recognition [16]–[19]. Such studies show that significant progress has been made to comprehend face perception in general and establishing brain regions responsible for face perception. Further, researchers have studied the effect of ancillary information on face recognition [20]. However, to the best of our knowledge, this research is the first work which attempts to understand the functioning of the brain for processing genuine and imposter pairs, i.e. *face verification task*, given familiar and unfamiliar faces. *Decoding* regions of activations and corresponding models for such a task can help face recognition researchers in developing improved algorithms. This paper focuses on this particular aspect of face verification and presents the following contributions:

- Design cognitive neuroscience experiments to observe brain activations while the participant is performing a face verification task. This is the first cognitive neuroscience experimental design for the same. While designing the experiment, special attention has been paid to suppress activations by memory recall.
- Propose a novel machine learning based approach to predict if the stimuli observed are genuine or imposter using fMRI data only. This approach is the first step in designing a brain-like face verification algorithm.

[1]Given a pair of face stimulus images, if both images correspond to the same subject, the pair is considered genuine; otherwise it is considered as imposter.

978-1-5090-6182-2/17/$31.00 ©2017 IEEE

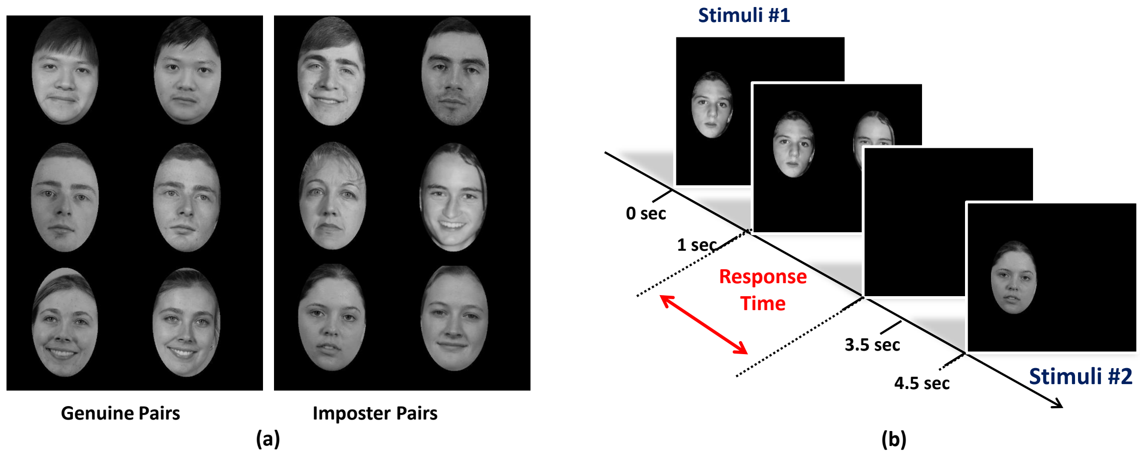

Fig. 1: (a) Sample stimuli for genuine and imposter pairs; (b) protocol for face verification task. The response time is marked in red and the maximum limit on response time is 2.5 seconds (red arrow). The screen turns blank after the participant provides response till 4.5 seconds before the next stimulus is shown.

II. Neurocognitive Study

Individuals perform face recognition innumerable times during their daily life. In most cases, it is an effortless task which is performed with utmost ease. However, researchers are yet to completely understand the intricacies of neural processing when a face is seen. This research aims to investigate the neural correlates of face verification (by one-to-one matching) using functional magnetic resonance imaging (fMRI). fMRI is a non-invasive neuroimaging technique which uses the relationship between brain activity and the local cerebral blood flow. Oxygen concentration in the blood is used as an indirect marker of the underlying neuronal activity in the brain; termed as the Blood Oxygenation Level Dependent (BOLD) signal. Researchers have been using fMRI scans and studying these BOLD signals to understand brain functioning [16], [18], [21], [22]. In this study, we try to decode the neuronal activity in the brain while performing face verification task (i.e. given a pair of face images, identify whether the pair is genuine or imposter). Participants are presented with a pair of face image stimulus and are required to decide whether the given pair of images belong to the same individual or not. Once the data is collected, responses are analyzed in terms of both cognitive or behavioral analysis and neuroimaging results. We further present the details about the data collection procedure, experimental protocol, and the results obtained.

A. Data Collection

Due to the unavailability of fMRI datasets in the literature which involve matching faces side-by-side, we collected data for the task of face verification. The data collection process involved the creation of task-specific stimuli, formulation of an experimental design, selection of participants and collection of fMRI scans, details of which are given below.

1) Experimental Design: The paradigm used in the fMRI experiment is an event-related face-matching task. Each stimuli consisted of gray scale images of face pairs displayed side by side where background noise (non-face area) is suppressed by applying an oval mask as shown in Fig. 1 (a). When viewing the stimulus, the participants were instructed to respond "Yes" if they thought the face-pair belongs to the same person and "No" otherwise, using a controller provided to them.

Each face-matching task is performed in a segment of 4.5 seconds where one stimulus pair is shown per segment. As shown in Fig. 1(b), in each segment, the first image of the pair is shown for the initial 1 second, followed by both the images being displayed, for a maximum of 2.5 seconds. Participants responded within these 2.5 seconds. As soon as a response is recorded, a blank black screen is displayed for the remaining 4.5 seconds segment and is treated as the Inter-Stimuli Interval (ISI). This functions as the resting phase (baseline) between each stimuli.

The entire experiment consisted of 4 fMRI runs with 60 segments in each run, making a total of 240 face-pair stimuli. No face-pair stimuli were overlapping across the different runs. Out of the 240 stimuli, 135 pairs are genuine (match pairs) and 105 pairs are imposter (non-match pairs). Face-pair stimuli were created using the images from popular face databases.

2) Participants: The study is performed on a total of 10 healthy subjects (5 males and 5 females) in the age group of 20-27 years. The subjects were briefed on the experimental design, type of stimuli, and the task before the data collection process. All subjects provided written consent to be a part of the study. The study has been approved by IIIT-Delhi Ethics board.

3) fMRI Data Acquisition: The imaging sequence is an interleaved T2*-weighted gradient echo sequence (from negative to positive direction) with 35 axial slices (slice thickness =

3.5 mm, slice spacing = 0.0 mm, repetition time (TR) = 2.0 seconds, echo time (TE) = 30 ms, flip angle = 65^0, field of view (FOV) = 224 mm, matrix = 64×64) and 128 volumes are captured per run (each volume is captured in 2 seconds with no gap of time between volumes) on a 3T GE Machine.

Anatomical structural scan for each subject is acquired using structural MRI scans. These scans are T1-weighted sequence with 172 sagittal slices in interleaved sequence (slice thickness = 1 mm, TR = 600 ms, flip angle = 10^0, field of view = 224 mm, matrix = 256×256).

B. Data Analysis and Results

In this section, the genuine-imposter responses provided by the participants during the fMRI task are analyzed. As explained above, the experimental protocol required the participants to perform face verification, i.e. verify whether a given stimulus, belonged to the same individual or not. With 240 stimuli and 10 participants, a total of 2400 responses are collected. We observe the correct verification accuracy[2] of **64.79**% across all participants. Upon further analysis, female participants correctly verified face stimuli with **67.17**% accuracy as compared to **62.41**% by male participants.

TABLE I: Confusion matrix for the behavioral responses for the face verification task. Performance for all participants, male participants, and female participants are reported separately.

			Response Label	
			Genuine	Imposter
All	Actual Label	Genuine	66.22%	33.56%
		Imposter	36.88%	62.95%
			Genuine	Imposter
Males	Actual Label	Genuine	63.26%	36.31%
		Imposter	38.79%	61.21 %
			Genuine	Imposter
Females	Actual Label	Genuine	63.55%	36.45%
		Imposter	27.27	72.32 %

Table I shows the confusion matrix in terms of actual label and response label for all 10 participants, as well as gender-wise results. It is important to note that due to some stimuli on which no response was given by the participant, the values across each row might not add up to a 100%. Intrigued by the variations in the true positive and true negative values from the confusion matrix, we further analyzed the performance of male and female participants on genuine and imposter pairs separately. For genuine pairs, male participants provided true positive accuracy of 63.26% and females provided 63.55% accuracy. However, it is interesting to observe that in case of imposter verification, females yield 72.32% and males yield 61.21% true negative rates, respectively.

[2]The accuracy is calculated as:

$$Accuracy = 100 \times \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives respectively.

To verify the statistical significance of the results, one-tailed z-test of proportions is applied with a significance level of 0.01 on the classification accuracy, TPR and TNR between males and females. A p-value of 0.0073 is obtained for accuracy comparison in favor of females performing better face verification as compared to males (67.17% over 62.41%). A one-tailed z-test of proportions on the true positive rates of both genders does not yield any statistical difference. However, a p-value of 0.0001 is observed for comparison of true negative rates of males (61.21%) and females (72.32%). This demonstrates the statistical significance of females performing better in determining imposters as compared to males.

The fMRI scans are also analyzed in terms of neural activations. fMRI scans require detailed pre-processing before they can be used to model the haemodynamic response. Preprocessing and analysis of the fMRI data was performed using the software Statistical Parametric Mapping (SPM 12; Wellcome Department of Cognitive Neurology) [23] with MATLAB R2016b. The pre-processing pipeline is as follows:

- Slice-time correction: All the functional volumes were slice time corrected to assign all the slices within a single volume to the same time point.
- Re-alignment: It was done to eliminate within-subject motion artifacts. The first image volume was used as the reference volume and all the other time series images within the subject were aligned with respect to the reference image using a least square minimization and a 6-parameter (rigid body) spatial transformation.
- Co-registration: Generally, structural scans are high resolution scans of the brain which provide a detailed structure of the organ. As compared to structural scans, functional scans do not have as much information. In this step, to maximize the information about the structure of the brain, the functional scans are projected onto the structural scans.
- Segmentation: All images are segmented into gray matter and white matter. This is done so as to extract the relevant voxels for analysis.
- Normalization: Since the brain shape and size of each individual is different, the pre-processed images are then aligned to a standard template space so as to perform inter-subject registration of scans using the MNI (Montreal Neurological Template) [24] space.
- Smoothing: A Gaussian kernel with the width of 4mm is used to remove any noise present in the images.

After pre-processing the data, the BOLD signal is modeled and activations in the brain are observed. We performed linear contrast using General Linear Model (GLM) framework over the pre-processed data to produce Statistical Parametric Map (SPM) of the effect of matching two face images. For each subject, a first level model was created and patterns of significant activation associated with face verification were identified by appropriately weighting the estimated model using simple T-contrasts (face verification > baseline) and statistical parametric maps (t-maps) and contrast images were

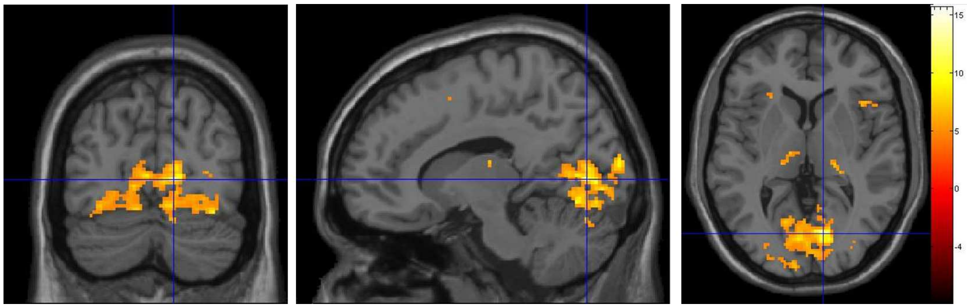

Fig. 2: Activations observed for the task of face verification at $p \leq 0.001$. The cross-hair points at the global maxima $(x = 12, y = -74, z = 4)$.

TABLE II: MNI co-ordinates (x, y, z) of the peak activations, observed during the task of face verification, over all the participants. z-scores corresponding to the activations for $p \leq 0.001$ (uncorrected) are also reported.

MNI coordinate of peak activation			Brain Region	z-score
x	y	z		
12	-74	4	Calcarine Sulcus (right)	5.26
10	-80	-10	Lingual Gyrus (right)	4.76
18	-16	14	Caudate (right)	4.69
-4	-2	64	Supplementary Motor Area (left)	4.69
-28	-48	20	Fusiform Gyrus (left)	4.57
-44	-14	54	Postcentral Gyrus (left)	4.41
-32	-16	-6	Putamen (left)	4.21
8	32	40	Superior Frontal Gyrus, Medial (left)	4.17
28	-52	46	Angular Gyrus (right)	4.16
-20	-32	0	Hippocampus (left)	4.11
-6	-16	-10	Thalamus (left)	3.99

generated for the whole brain. Group level analysis was performed by computing one-sample second level t-statistic (GLM random effects analysis) using the contrast images across all subjects and a statistical parametric map of the effect of matching two face images was generated, thresholded at $p \leq 0.001$.

Table II presents detailed coordinates, corresponding brain regions and z-scores and Fig. 2 depicts the activations observed at the given p-value at the global peak. The analysis revealed significant activations in the areas of calcarine sulcus, where the primary visual cortex is concentrated, lingual gyrus (part of occipital lobe), which is responsible for visual processing, and fusiform gyrus. These results agree with previous studies which have shown the existence of face-selective regions in the fusiform gyrus and anterior inferotemporal cortex [7], [25]. Significant activations are also observed in the left fusiform gyrus, which is responsible for detecting *face-like* features. Similar activations in the areas of hippocampus, caudate, postcentral gyrus, and superior frontal gyrus are observed during recognition of famous faces [26]. Activations in the areas of thalamus and angular gyrus assert that the participants were alert and attentive to the stimuli [27].

The observations obtained from the neuroimaging experiment are in accordance with existing literature. The activation of meaningful regions responsible for visual processing and face perception motivate us to further explore the information from the fMRI brain scans. The following section presents a novel learning based approach to predict the stimulus viewed by a participant for a given fMRI scan.

III. Neural Pattern based Face Stimuli Classification by Two-level FMRI Dictionary Pairs

Recently, machine learning algorithms have been developed to decode the cognitive state of the subject which are useful for various applications such as depression detection [28], emotion identification [29], word pronunciation [30], and lie detection [31]. By considering the neural activity measured at different locations, machine learning techniques help in discovering patterns for automated analysis and model design [32]. Several studies in the literature have also focused on exploring the areas of activation in the brain during the face perception task [33], [34]. On the other hand, the quantifiable localizations of patterns of neural activity have gained fairly less attention, particularly as a way of decoding the cognitive state.

In this section, the emphasis is on the *inverse* problem of classifying the stimulus for a given fMRI scan, i.e. predicting the stimulus as genuine or imposter face pair. It is our belief that there is an inherent difference in the brain's functional network while processing genuine face pairs as compared to imposter face pairs. Hence, the aim is to utilize machine learning techniques to design a framework which can decode whether the face pairs seen by a human subject are genuine or imposter. As shown in Fig. 3, the proposed framework involves selection of scans for stimuli pairs and regions of interest (ROIs) based mask generation. This is followed by face-specific and decision-specific ROI-based feature extraction. Dictionary pairs are trained for each ROI and combined to learn two-level dictionary pairs. This two-level dictionary framework is designed to learn features of each specific brain area as well as a cumulative representation of regions involved in face processing and decision making. The last step is decision-level fusion of outputs from face-specific and decision-specific fMRI dictionary pairs to obtain the final prediction. These steps are explained in detail below.

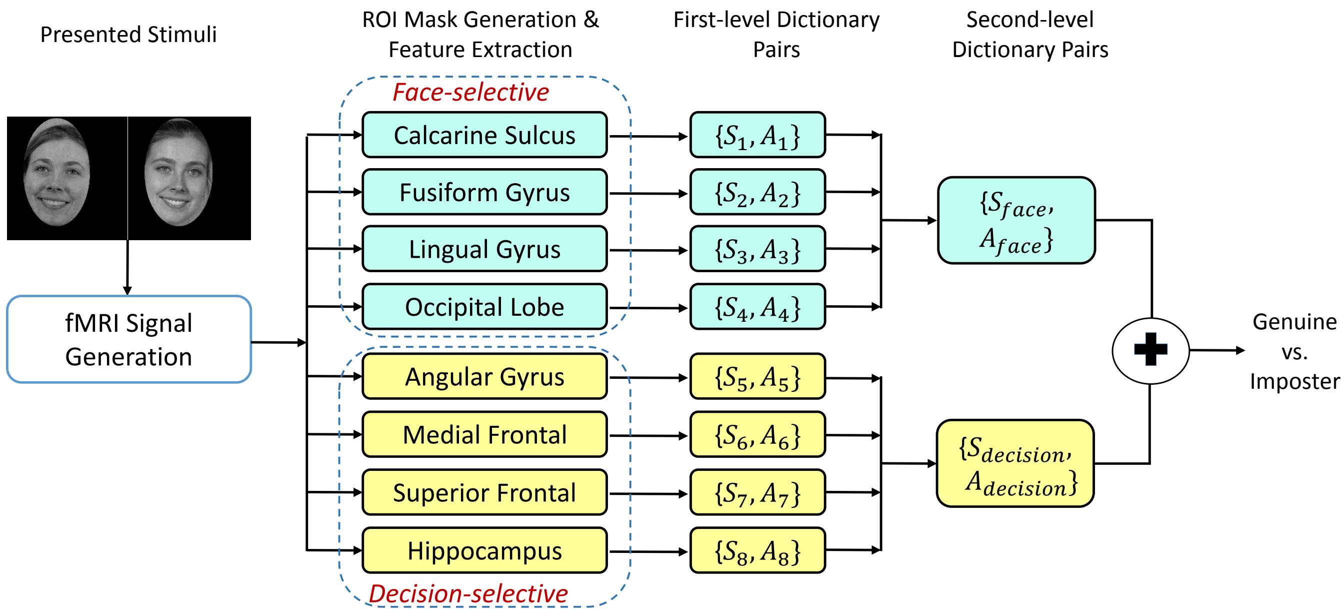

Fig. 3: Proposed two-level fMRI dictionary based framework which involves fusion of information from face-selective and decision-selective regions of the brain in hierarchical manner. In the first-level, ROI based synthesis and analysis dictionary pairs (S_i, A_i) are trained to model representations through sparse signals pertaining to the region. In the second-level, dictionaries are trained on the concatenated output from first-level to process the information from face-specific and decision-specific brain areas holistically.

A. Selection of fMRI Scans for Stimuli Pairs

As described in the previous section, fMRI responses are collected while the subject is viewing genuine or imposter face pair. The fMRI scans corresponding to the two types of stimuli are labeled and separated into the two classes for each participant. Due to the nature of images being presented sequentially, confounding scans that may have huge overlaps in the BOLD response with respect to different stimuli are removed.

B. Mask Generation of Face-selective and Decision-selective ROIs

The dimensionality of fMRI data is very high; for example, a scan with 35 slices of the human brain with a resolution of 64×64 during one time-stamp leads to a feature vector of size 143,360. Therefore, creating masks to reduce the dimensionality and redundancy is a crucial task before analyzing the fMRI data. Based on the type of stimulus presented to the participant, there are two different approaches for mask creation to classify a given fMRI scan. Both approaches have their respective advantages and disadvantages. In the first approach, the fMRI data from the entire brain (removing the non-brain regions) is taken into account for computational purposes [29], [31]. This approach accounts for less apriori knowledge regarding functionality of brain regions. However, the data collected from the brain is complex and takes into account multiple interactions of different regions. Also, it often leads to overfitting due to the high dimensionality of the data. In the second approach, analysis is restricted to specific Regions of Interest (ROIs) chosen from the literature that reduces the dimensions of the data and may lead to better discrimination [35], [36]. Information from each region of interest is computed to form local features, which are robust to inter-subject functional distinctions within the brain.

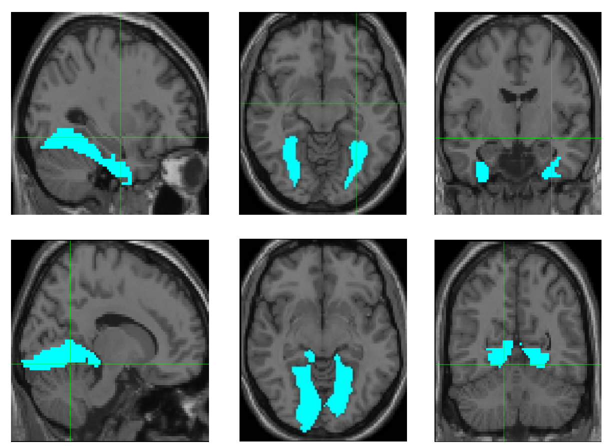

Fig. 4: Fusiform Gyrus (row 1) and Lingual Gyrus (row 2) masks generated for the Regions of Interest (ROIs). The underlying anatomical brain image is for representation purposes only.

In this paper, a fusion of ROI-based features is proposed for combining different information sources to yield better discriminability for classifying stimulus pair as genuine or imposter, given an fMRI scan. Based on previous studies, two groups of ROIs are chosen. Face-specific brain regions are related to face perception and processing. In this group, the included brain areas are calcarine sulcus [37], fusiform gyrus [33], [38], lingual gyrus [39], [40], and occipital lobe [41]. The fusiform gyrus contains the fusiform face area that is associated with face perception and face recognition while the lingual gyrus is associated with complex visual processing [42]. Likewise, regions involved in decision making process (decision-specific) are selected to form the second group. ROIs included in this group are angular gyrus [43], hippocampus [44], medial frontal [45], and superior frontal gyrus [46]. These regions are also observed in the neural activations seen in Table II which further strengthens the motivation for selection of these areas. In order to extract these ROIs from each scan, anatomical masks are created using the Automated Anatomical Labeling (AAL) atlas [47]. They are resized and updated according to the data obtained from the ten subjects of the study. Fig. 4 illustrates the sample anatomical masks generated for two ROIs: fusiform gyrus and lingual gyrus.

C. Encoding fMRI signals using Two-level Dictionary Pairs

Neural representations are extracted for face-selective and decision-selective ROIs by applying the anatomical masks computed in the previous step. It has been found that sparse neurons do not respond to the input independently and encode specific concepts together [48]. Thus, learning dictionaries that encode a sparse basis representation of the input data, are a natural way to represent fMRI signals. Vast literature is available that utilizes sparse coding and dictionary learning to analyze the fMRI activations [49], [50]. In this paper, an fMRI dictionary pair framework is utilized to learn the intrinsic properties of the fMRI signals and encode them according to the external stimuli shown and decision made by the participant.

Gu et al. [51] introduced the Dictionary Pair Learning (DPL) framework, where two dictionaries are jointly trained to learn representations through linear projection. These two dictionaries are termed as analysis and synthesis dictionary and together encode the discriminative information present in the input vector and its reconstruction. The model can be described as:

$$\{\boldsymbol{\mathcal{A}}^*, \boldsymbol{\mathcal{S}}^*\} = \underset{\boldsymbol{\mathcal{A}},\boldsymbol{\mathcal{S}}}{\operatorname{argmin}} \sum_{k=1}^{K} || \, \boldsymbol{\mathcal{X}}_k - \boldsymbol{\mathcal{S}}_k \boldsymbol{\mathcal{A}}_k \boldsymbol{\mathcal{X}}_k \, ||_F^2 + \gamma \, || \, \boldsymbol{\mathcal{A}}_k \bar{\boldsymbol{\mathcal{X}}}_k \, ||_F^2, \quad s.t \, || \, \boldsymbol{d}_i \, ||_2^2 \leq 1 \tag{2}$$

where, $\boldsymbol{\mathcal{S}}$ represents the synthesis dictionary used to reconstruct the input matrix $\boldsymbol{\mathcal{X}}$; $\boldsymbol{\mathcal{A}}$ represents the analysis dictionary used to encode input $\boldsymbol{\mathcal{X}}$; $\boldsymbol{\mathcal{A}_k}$ and $\boldsymbol{\mathcal{S}_k}$ represent the sub-dictionary pair corresponding to class k; $\bar{\boldsymbol{\mathcal{X}}}_{\boldsymbol{k}}$ represents the complementary data matrix of $\boldsymbol{\mathcal{X}}_k$ in the training set; γ>0 is a scalar constant that denotes the regularization parameter to control the discriminative property of $\boldsymbol{\mathcal{A}}$, and $\boldsymbol{d}_i$ denotes the ith item of synthesis dictionary $\boldsymbol{\mathcal{S}}$. The role of the analysis dictionary $\boldsymbol{\mathcal{A}}$ is to help in discrimination, where the sub-dictionary $\boldsymbol{\mathcal{A}_k}$ can project the samples from class $i, i \neq k$ to 0. The role of the synthesis dictionary $\boldsymbol{\mathcal{S}}$ is to minimize the reconstruction error. The above framework enforces group sparsity on the input matrix using the analysis dictionary $\mathcal{A}$ but does not utilize the time consuming l_0 or l_1 norm computation.

To learn a discriminatory model for distinguishing between genuine and imposter stimuli using fMRI scans, fMRI dictionary pairs S_i, A_i (where $i = 1 \ldots 8$) are learned for each ROI. These first-level dictionaries encode the individual ROI-specific features. The reconstructed outputs (R_i) from the learned dictionaries are concatenated to form inputs to face-specific and decision-specific fMRI dictionary pairs. Therefore, the input to second-level face-specific dictionary (D_{face}) is $[R_1, R_2, R_3, R_4]$ corresponding to calcarine sulcus, fusiform gyrus, lingual gyrus, and occipital lobe. Likewise, the input to second-level decision-specific dictionary ($D_{decision}$) is $[R_5, R_6, R_7, R_8]$ corresponding to angular gyrus, medial frontal gyrus, superior frontal gyrus, and hippocampus. These second-level dictionaries learn cumulative features of face-specific and decision-specific brain areas in a hierarchical fashion. The classification outputs from the second-level face-specific dictionary pair (D_{face}) is labeled as Y_{face} while second-level decision-specific dictionary pair ($D_{decision}$) is labeled as $Y_{decision}$.

D. Decision-level Fusion of Second-level fMRI Dictionary Pairs

In the proposed framework, decision-level fusion is utilized to combine the dictionary pair classification outputs Y_{face} and $Y_{decision}$. For integrating the information represented by the learned dictionary pairs, the final classification, Y_{output} of an fMRI scan into genuine or imposter stimuli is performed by applying a logical AND on the individual decisions.

$$Y_{output} = Y_{face} \wedge Y_{decision} \tag{3}$$

IV. Experiments and Analysis

In this experiment, unseen training and testing partitioning is performed with five-fold subject-based cross validation. Hence, in each fold, scans of eight subjects are used for training and testing is performed on the remaining two subjects. Eight anatomical masks corresponding to the face-selective ROIs (R_1: calcarine sulcus, R_2: fusiform gyrus, R_3: lingual gyrus, and R_4: occipital lobe) and decision-selective ROIs (R_5: angular gyrus, R_6: hippocampus, R_7: medial frontal gyrus, and R_8: superior frontal gyrus) are computed followed by feature extraction from these ROIs. fMRI dictionary pairs are learned for each of these regions. The reconstructed samples using the dictionary pairs from face-selective ROIs are concatenated to form dictionary pair D_{face} to yield genuine vs imposter pair classification. The same step is followed with decision-selective ROIs to generate the dictionary pair $D_{decision}$. The predictions from the two fMRI dictionary pairs are combined at the decision-level to obtain the final classification output. The experiments are run on a Linux machine with Intel Core i7-4770 CPU @ 3.40GHz and 32 GB memory. The key findings from the experimental results are explained below.

TABLE III: Stimuli classification accuracy using decision-level fusion of ROI-based machine learning algorithms.

Machine Learning Technique	Accuracy (%)
Linear Discriminant Analysis	48.48
Decision Trees	49.27
Neural Network	51.30
Naive Bayes	54.82
Proposed Two-level Dictionary	**59.35**

- The average five-fold classification accuracy obtained using the proposed two-level fMRI dictionary pair framework is **59.35%** which is $\approx$ 4% less than classification performance by human subjects. This indicates the existence of separable and discriminatory cognitive states corresponding to the type of face stimuli presented to the participants.
- To demonstrate the effectiveness of the proposed framework for this problem, comparison has been drawn with other commonly used machine techniques such as Naive Bayes, Linear Discriminant Analysis (LDA), Decision Tree, and Neural Network as tabulated in Table III. It is observed that the proposed two-level fMRI dictionary learned framework outperforms other techniques by at least 4%. The lower classification accuracies obtained by other machine learning techniques illustrate the challenging nature of the classification problem.
- The training time for the proposed two-level fMRI dictionary pair framework (consisting of ten dictionary pairs) is also analyzed. In total, the first-level dictionary pairs training took 1.04 seconds as compared to 1.45 seconds by the second-level dictionary pairs training.
- Comparative analysis is performed with one-level fMRI dictionary pairs computed from the features of the whole brain. The classification accuracy with this approach is 52.80%. This demonstrates the efficacy of a region-wise approach over less discriminatory features from the full brain region.
- Region-specific analysis is also performed to gain better insight of the obtained results. When classification is performed by dictionary pairs of face-selective and decision-selective ROIs separately, classification accuracies of 57.60% and 53.98% are observed, respectively. These results suggest that (i) the effect of individual ROIs might not be sufficient to understand the complex functionality of the brain and (ii) activations corresponding to the different brain areas need to be modeled to understand the functionality of the brain as a whole.

V. Conclusion

The visual processing performed by the human brain is inherently complex which has intrigued the research community. The fusiform face area (FFA) is known to be involved in face perception and recognition. At the same time, researchers have successfully developed automatic face verification algorithms that can match face images. In this paper, we bridge the gap between both of these areas by exploring the process of face verification through fMRI data analysis. A total of 2400 fMRI responses with genuine and imposter stimuli pairs for a novel face verification task are collected from 10 participants. The neural activations due to the face verification task are analyzed. Brain areas belonging to visual and face processing such as fusiform gyrus, lingual gyrus and calcarine are significantly activated at $p \leq 0.001$. It is also observed that female participants are able to identify imposter pairs significantly better than male participants. Based on the neuroimaging results, a novel two-level fMRI dictionary learning based framework with face-specific and decision-specific regions of interest is proposed to predict the type of stimuli (genuine or imposter) shown to the participants during the fMRI scans. Classification accuracy of 59.35% is observed using the proposed framework and comparative analysis is also performed with other machine learning techniques. We observe that the difference in the classification accuracy of humans and the proposed framework is only 4%. It is our belief that understanding the human visual system will play a crucial role in creating stable and accurate automatic face verification models, and this study is the first step towards that direction. In future, we aim to extend the neuroimaging analysis using fMRI activations obtained with corrected thresholds and further incorporate sophisticated techniques for noise removal from the signals.

Acknowledgment

This research is supported in part from a grant from Ministry of Electronics and Information Technology, India. The authors acknowledge the feedback provided by the reviewers.

References

[1] I. J. Cox, J. Ghosn, and P. Yianilos, "Feature-based face recognition using mixture-distance," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1996, pp. 209–216.

[2] G. Goswami, R. Bhardwaj, R. Singh, and M. Vatsa, "MDLFace: Memorability augmented deep learning for video face recognition," in *IEEE International Joint Conference on Biometrics*, 2014, pp. 1–7.

[3] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 815–823.

[4] R. Singh, M. Vatsa, and A. Noore, "Face recognition with disguise and single gallery images," *Image and Vision Computing*, vol. 27, no. 3, pp. 245–257, 2009.

[5] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701–1708.

[6] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, Jul 1997.

[7] N. Kanwisher, J. McDermott, and M. M. Chun, "The fusiform face area: A module in human extrastriate cortex specialized for face perception," *The Journal of Neuroscience*, vol. 17, no. 11, pp. 4302–4311, 1997.

[8] E. McKone, K. Crookes, and N. Kanwisher, "The cognitive and neural development of face recognition in humans," *The Cognitive Neurosciences*, vol. 4, pp. 467–482, 2009.

[9] J. A. Collins and I. R. Olson, "Beyond the FFA: The role of the ventral anterior temporal lobes in face processing," *Neuropsychologia*, vol. 61, pp. 65 – 79, 2014.

[10] D. Pitcher, V. Walsh, and B. Duchaine, "The role of the occipital face area in the cortical face perception network," *Experimental Brain Research*, vol. 209, no. 4, pp. 481–493, 2011.

[11] M. I. Gobbini and J. V. Haxby, "Neural response to the visual familiarity of faces," *Brain Research Bulletin*, vol. 71, pp. 76 – 82, 2006.

[12] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.

[13] N. Kohli, M. Vatsa, R. Singh, A. Noore, and A. Majumdar, "Hierarchical representation learning for kinship verification," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 289–302, 2017.

[14] H. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Recognizing surgically altered face images using multiobjective evolutionary algorithm," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 89–100, 2013.

[15] N. Kohli, D. Yadav, and A. Noore, "Multiple projective dictionary learning to detect plastic surgery for face verification," *IEEE Access*, vol. 3, pp. 2572–2580, 2015.

[16] J. T. Kaplan, L. Aziz-Zadeh, L. Q. Uddin, and M. Iacoboni, "The self across the senses: an fMRI study of self-face and self-voice recognition," *Social Cognitive and Affective Neuroscience*, vol. 3, no. 3, pp. 218–223, 2008.

[17] V. S. Natu and A. J. O'Toole, "Spatiotemporal changes in neural response patterns to faces varying in visual familiarity," *NeuroImage*, vol. 108, pp. 151–159, 2015.

[18] S. M. Platek and S. M. Kemp, "Is family special to the brain? an event-related fMRI study of familiar, familial, and self-face recognition," *Neuropsychologia*, vol. 47, no. 3, pp. 849 – 858, 2009.

[19] L. Q. Uddin, J. T. Kaplan, I. Molnar-Szakacs, E. Zaidel, and M. Iacoboni, "Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: an event-related fMRI study," *NeuroImage*, vol. 25, no. 3, pp. 926 – 935, 2005.

[20] U. Toseeb, D. R. T. Keeble, and E. J. Bryant, "The significance of hair for face recognition," *PLOS ONE*, vol. 7, no. 3, pp. 1–8, 2012.

[21] J. Liu, A. Harris, and N. Kanwisher, "Perception of face parts and face configurations: an fMRI study," *Journal of Cognitive Neuroscience*, vol. 22, no. 1, pp. 203–211, 2010.

[22] J. Parvizi, C. Jacques, B. L. Foster, N. Withoft, V. Rangarajan, K. S. Weiner, and K. Grill-Spector, "Electrical stimulation of human fusiform face-selective regions distorts face perception," *The Journal of Neuroscience*, vol. 32, no. 43, pp. 14 915–14 920, 2012.

[23] "Statistical parametric mapping toolbox," http://www.fil.ion.ucl.ac.uk/spm/, [Online; accessed 1-December-2016].

[24] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, and D. L. Collins, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage*, vol. 54, no. 1, pp. 313 – 327, 2011.

[25] I. Gauthier, M. J. Tarr, J. Moylan, P. Skudlarski, J. C. Gore, and A. W. Anderson, "The fusiform "face area" is part of a network that processes faces at the individual level." *Journal Cognitive Neuroscience*, vol. 12, pp. 495–504, 2000.

[26] C. Elfgren, D. Westen, U. Passant, E.-M. Larsson, P. Mannfolk, and P. Fransson, "fMRI activity in the medial temporal lobe during famous face processing," *NeuroImage*, vol. 30, no. 2, pp. 609 – 616, 2006.

[27] P. Bourke, S. Brown, E. Ngan, and M. Liotti, "Functional brain organization of preparatory attentional control in visual search," *Brain Research*, vol. 1530, pp. 32 – 43, 2013.

[28] L. Oliveira, C. D. Ladouceur, M. L. Phillips, M. Brammer, and J. Mourao-Miranda, "What does brain response to neutral faces tell us about major depression? evidence from machine learning and fmri," *PLOS ONE*, vol. 8, no. 4, pp. 1–8, 2013.

[29] K. S. Kassam, A. R. Markey, V. L. Cherkassky, G. Loewenstein, and M. A. Just, "Identifying emotions on the basis of neural activation," *PLOS ONE*, vol. 8, no. 6, pp. 1–12, 2013.

[30] T. Bitan, A. Frid, H. Hazan, L. M. Manevitz, H. Shalelashvili, and Y. Weiss, "Classification from generation: Recognizing deep grammatical information during reading from rapid event-related fmri," in *International Joint Conference on Neural Networks*, 2016, pp. 4637–4642.

[31] C. Davatzikos, K. Ruparel, Y. Fan, D. Shen, M. Acharyya, J. Loughead, R. Gur, and D. D. Langleben, "Classifying spatial patterns of brain activity with machine learning methods: application to lie detection," *Neuroimage*, vol. 28, no. 3, pp. 663–668, 2005.

[32] Y. Ni, C. Chu, C. J. Saunders, and J. Ashburner, "Kernel methods for fMRI pattern prediction," in *IEEE International Joint Conference on Neural Networks*, 2008, pp. 692–697.

[33] N. George, R. J. Dolan, G. R. Fink, G. C. Baylis, C. Russell, and J. Driver, "Contrast polarity and face recognition in the human fusiform gyrus," *Nature Neuroscience*, vol. 2, no. 6, pp. 574–580, 1999.

[34] G. McCarthy, A. Puce, J. C. Gore, and T. Allison, "Face-specific processing in the human fusiform gyrus," *Journal of Cognitive Neuroscience*, vol. 9, no. 5, pp. 605–610, 1997.

[35] V. Natu and A. J. O'Toole, "The neural processing of familiar and unfamiliar faces: A review and synopsis," *British Journal of Psychology*, vol. 102, no. 4, pp. 726–747, 2011.

[36] V. S. Natu, F. Jiang, A. Narvekar, S. Keshvari, V. Blanz, and A. J. O'Toole, "Dissociable neural patterns of facial identity across changes in viewpoint," *Journal of Cognitive Neuroscience*, vol. 22, no. 7, pp. 1570–1582, 2010.

[37] J. V. Haxby, L. G. Ungerleider, B. Horwitz, J. M. Maisog, S. I. Rapoport, and C. L. Grady, "Face encoding and recognition in the human brain," *Proceedings of the National Academy of Sciences*, vol. 93, no. 2, pp. 922–927, 1996.

[38] B. Rossion, L. Dricot, A. Devolder, J.-M. Bodart, M. Crommelinck, B. Gelder, and R. Zoontjes, "Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus," *Journal of Cognitive Neuroscience*, vol. 12, no. 5, pp. 793–802, 2000.

[39] M. L. Kesler, A. H. Andersen, C. D. Smith, M. J. Avison, C. E. Davis, R. J. Kryscio, L. X. Blonder *et al.*, "Neural substrates of facial emotion processing using fMRI," *Cognitive Brain Research*, vol. 11, no. 2, pp. 213–226, 2001.

[40] A. Puce, T. Allison, J. C. Gore, and G. McCarthy, "Face-sensitive regions in human extrastriate cortex studied by functional MRI," *Journal of Neurophysiology*, vol. 74, no. 3, pp. 1192–1199, 1995.

[41] J. K. Steeves, J. C. Culham, B. C. Duchaine, C. C. Pratesi, K. F. Valyear, I. Schindler, G. K. Humphrey, A. D. Milner, and M. A. Goodale, "The fusiform face area is not sufficient for face recognition: evidence from a patient with dense prosopagnosia and no occipital face area," *Neuropsychologia*, vol. 44, no. 4, pp. 594–609, 2006.

[42] E. Macaluso, C. D. Frith, and J. Driver, "Modulation of human visual cortex by crossmodal spatial attention," *Science*, vol. 289, no. 5482, pp. 1206–1208, 2000.

[43] R. H. Grabner, D. Ansari, K. Koschutnig, G. Reishofer, F. Ebner, and C. Neuper, "To retrieve or to calculate? left angular gyrus mediates the retrieval of arithmetic facts during problem solving," *Neuropsychologia*, vol. 47, no. 2, pp. 604–608, 2009.

[44] E. Tulving and H. J. Markowitsch, "Episodic and declarative memory: role of the hippocampus," *Hippocampus*, vol. 8, no. 3, pp. 198–204, 1998.

[45] A. Talati and J. Hirsch, "Functional specialization within the medial frontal gyrus for perceptual go/no-go decisions based on "what", "when" and "where" related information: an fMRI study," *Journal of cognitive neuroscience*, vol. 17, no. 7, pp. 981–993, 2005.

[46] F. Du Boisgueheneuc, R. Levy, E. Volle, M. Seassau, H. Duffau, S. Kinkingnehun, Y. Samson, S. Zhang, and B. Dubois, "Functions of the left superior frontal gyrus in humans: a lesion study," *Brain*, vol. 129, no. 12, pp. 3315–3328, 2006.

[47] T. Nathalie, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.

[48] I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'ardenne, W. Richter, J. Cohen, and J. Haxby, "Independent component analysis for brain fMRI does not select for independence," *Proceedings of the National Academy of Sciences*, vol. 106, no. 26, pp. 10 415–10 422, 2009.

[49] K. Lee, S. Tak, and J. C. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1076–1089, 2011.

[50] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Fast and incoherent dictionary learning algorithms with application to fMRI," *Signal, Image and Video Processing*, vol. 9, no. 1, pp. 147–158, 2015.

[51] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Advances in Neural Information Processing Systems*, 2014, pp. 793–801.