Journal of Human Performance in Extreme Environments

Volume 14 | Issue 1 Article 3

Published online: 1-24-2018

Deep Gaze Velocity Analysis During Mammographic Reading for Biometric Identification of Radiologists

Hong-Jun Yoon
Biomedical Sciences, Engineering, and Computing Group Health Data Sciences Institute, Oak Ridge National Laboratory, yoonh@ornl.gov

Folami Alamudun
Biomedical Sciences, Engineering, and Computing Group Health Data Sciences Institute, Oak Ridge National Laboratory

Kathy Hudson
Department of Radiology, University of Tennessee Graduate School of Medicine

Garnetta Morin-Ducote
Department of Radiology, University of Tennessee Graduate School of Medicine

Georgia Tourassi
Biomedical Sciences, Engineering, and Computing Group Health Data Sciences Institute, Oak Ridge National Laboratory, tourassig@ornl.gov

Deep Gaze Velocity Analysis During Mammographic Reading for Biometric Identification of Radiologists

Cover Page Footnote

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. The study was supported by the Laboratory Directed Research and Development (LDRD) program of Oak Ridge National Laboratory, under LDRD project No. 6268.

Deep Gaze Velocity Analysis During Mammographic Reading for Biometric Identification of Radiologists

Hong-Jun Yoon and Folami Alamundun

Oak Ridge National Laboratory

Kathy Hudson and Garnetta Morin-Ducote

University of Tennessee Graduate School of Medicine

Georgia Tourassi

Oak Ridge National Laboratory

Abstract

Several studies have confirmed that the gaze velocity of the human eye can be utilized as a behavioral biometric or personalized biomarker. In this study, we leverage the local feature representation capacity of convolutional neural networks (CNNs) for eye gaze velocity analysis as the basis for biometric identification of radiologists performing breast cancer screening. Using gaze data collected from 10 radiologists reading 100 mammograms of various diagnoses, we compared the performance of a CNN-based classification algorithm with two deep learning classifiers, deep neural network and deep belief network, and a previously presented hidden Markov model classifier. The study showed that the CNN classifier is superior compared to alternative classification methods based on macro F_1-scores derived from 10-fold cross-validation experiments. Our results further support the efficacy of eye gaze velocity as a biometric identifier of medical imaging experts.

Keywords: convolutional neural networks, deep learning, eye tracking, gaze velocity

Introduction

Eye tracking has been studied extensively in its application as a biometric for the identification and authentication of individuals (Bednarik, Kinnunen, Mihaila, & Fränti, 2005; Deravi & Guness, 2011; Galdi, Nappi, Riccio, Cantoni, & Porta, 2013; Holland & Komogortsev, 2012; Maltoni & Jain, 2004; Rigas, Komogortsev, & Shadmehr, 2016; Yoon, Carmichael, & Tourassi, 2014). Findings from these studies suggest eye tracking not only provides a convenient way to capture "soft biometric" data (Galdi et al., 2013) but also an effective way of capturing physiological and behavioral aspects of brain-driven visuo-cognitive activity, both of which are less susceptible to falsification (Holland & Komogortsev, 2012; Rigas et al., 2016).

Recent advances in eye tracking device technology have enabled researchers to capture various eye-movement characteristics and explore the efficacy of these characteristics as biometric identifiers. For example, gaze trajectory (Deravi & Guness, 2011; Galdi et al., 2013), gaze velocity (Yoon et al., 2014), and pupillary characteristics (Bednarik et al., 2005) have been applied with reasonable success for biometric identification.

Kasprowski and Ober (2004) utilized a combination of eye reaction time and stabilization time as features to build a predictive model for biometric identification. They applied 10-fold cross-validation methods to test four predictive models (*k*-nearest neighbors, naive Bayes, C4.5 decision tree, and support vector machines) on eye tracking data collected from nine participants. They reported the highest average false acceptance rate of 1.48 achieved with *k*-nearest neighbors.

Galdi et al. (2013) developed a gaze analysis-based soft-biometric (GAS), predicated on stimulus-based viewing behavior (such as viewing behavior while observing a facial image). The GAS system used a fixed region of interest-based feature vector, which was computed using order-independent cumulative duration of fixations on the respective regions of interest. Subsequent test samples were identified using a Euclidean distance metric from user-dependent profiles.

http://dx.doi.org/10.7771/2327-2937.1088

Holland and Komogortsev (2012) evaluated eye movement-based measures as features for biometric identification. They recorded the eye movements of 32 participants (26 male and 6 female) using a head-mounted eye-tracking device. They extracted basic eye movement features and scanpath measures including: fixation count, fixation duration, saccade amplitude and velocity. Applying an information fusion method, they combined these features and reported a 27% error rate for personal identification.

Fookes, Maeder, Sridharan, and Mamic (2009) developed methods to characterize visual attention as a biometric for authentication and identification of image viewers. To characterize spatial and temporal patterns of visual attention, they proposed three techniques: principal component analysis ("eigenGaze"), linear discriminant analysis, and a fusion of distance measures. Their experimental results suggest that all three techniques can provide simple and effective biometrics for classifying a small database of individuals.

Previous studies have investigated more sophisticated techniques for characterization of gaze behavior. For example, Rothkopf and Pelz (2004) adapted hidden Markov models (HMMs) to characterize gaze velocities for the classification of different types of visual behavior. In our previous studies (Yoon et al., 2014, 2015), we also investigated the properties of gaze velocity using HMMs for general viewing of cognitive-dot stimuli related to the Gestalt grouping principles of similarity, continuation, proximity, and closure (Yoon et al., 2014), and for expert viewing of regions of interest within mammographic images (Yoon et al., 2015). Both studies suggested that gaze velocity is a promising biometric feature for general and medical viewing tasks. However, HMMs are probabilistically modelled and fail to capture features corresponding to patterns of eye movement that uniquely characterize the visual perception process of an individual. Investigating such characteristics requires more advanced feature engineering, which is costly and time consuming.

Deep learning has recently emerged as a highly effective machine learning approach capable of achieving high levels of data abstraction without requiring manual feature engineering (LeCun, Bengio, & Hinton, 2015). Convolutional neural networks (CNNs), a popular type of deep learning, are 'by design" capable of automatic feature representation by training multiple layers of convolutional filters. The classification capability of CNNs was demonstrated on several image classification tasks (Krizhevsky, Sutskever, & Hinton, 2012). Most studies utilized the CNNs for two-dimensional image classification tasks, though it is trivial to apply one-dimensional temporal sequence data just by aligning one-dimensional convolutional filters in temporal sequences.

In this work, we leverage the automatic feature representation advantage of deep learning for a gaze velocity-based biometric identification of medical experts. Specifically, we compare the performance of three deep learning methods (deep neural networks (DNNs), deep belief networks (DBNs), and CNNs) for personal identification of radiologists performing breast cancer screening tasks under real clinical conditions. Our overarching goal is to investigate the effectiveness of deep learning methods in extracting gaze velocity patterns which could serve as individual biometric identifiers.

In the following sections, we provide a brief literature review on the application of gaze velocity as a behavioral biometric. We describe the data collection protocol and the deep learning algorithms implemented for data analysis and provide experimental results. Finally, we discuss the study findings and give direction for future work.

Gaze Velocity as a Behavioral Biometric

A large proportion of eye movements are a physiological response to visual stimuli, such as reading texts, viewing pictures, or tracing targets. Previous studies have examined eye movements such as gaze location, fixations, pupil diameter, and other similar metrics to characterize visual behavior.

Henriksson, Pyykko, Schalen, and Wennmo (1980) presented several key findings on saccadic movement and velocity, including the considerable inter-subject variations in saccadic velocity. This finding was also confirmed by Schmidt, Abel, DellOsso, and Daroff (1979), who reported an associated inter-subject variability in peak velocity and amplitude.

Bednarik et al. (2005) developed eye movement-based biometric features from changes in pupil diameter, gaze velocity, and eye distance from 12 study participants while viewing a combination of stationary and moving objects. By applying dimensionality reduction of features based on Fourier transform followed by principal component analysis, they reported a 60% identification rate using a simple *k*-means clustering algorithm.

Silver and Biggs (2006) investigated a combination of keystroke and eye movement data for biometric identification using a reading-while-typing stimulus. They trained a probabilistic neural network on data from 21 study participants and reported an average accuracy of 96.6%. Although they reported superior performance from keystroke-based biometric features, they also noted that eye movement-based features showed promising results, which warrant further investigation.

Kinnunen, Sedlak, and Bednarik (2010) developed a task-dependent person identification system by applying Gaussian mixture models on feature vectors of short-term eye gaze direction. Their results suggested that there are task-dependent person-specific features in the eye movement, which may be useful in user authentication systems.

Cuong, Dinh, and Ho (2012) proposed a novel method for extracting eye movement features using mel-frequency cepstal coefficients. They reported an average identification rate of 92.35% on two open datasets provided by the

First Eye Movement Verification and Identification Competition (EMVIC 2012) using decision tree, Bayesian network, and support vector machine classifiers.

Zhang and Juhola (2012) explored properties of saccadic eye movements using a stimulus of a horizontal jumping point of light. Their methodology involved the application of machine learning techniques on features of saccadic amplitude, accuracy, latency, velocity, and acceleration. The tested techniques included multilayer perceptron networks, radial basis function networks, support vector machines, and logistic discriminant analysis. The experimental database was formed from 132 study participants, and the reported best verification accuracy was 89%.

In this study, we investigated local features of gaze velocity, which may not be captured effectively by aggregate measures other studies have applied. However, instead of spending effort on the manual curation of gaze velocity feature representations, we applied deep learning. Particularly, the CNN is a well-known automatic feature learner capable of dealing well with shift-variance. Thus, CNNs may represent better features of gaze velocity during fixation and saccadic movements of eyes. Since gaze behavior is task-dependent, we focus on a visual task that is known to be one of the most challenging in the medical imaging community, namely breast cancer screening using mammograms (Tourassi, 2005). Furthermore, several studies have shown that the radiologists' perceptual behavior changes over time as a function of clinical training and experience (Tourassi, Voisin, Paquit, & Krupinski, 2013; Voisin, 2013a,b). Therefore, deep learning presents a unique opportunity for data-driven modeling of individual perceptual behavior in the medical imaging domain without requiring time-consuming feature engineering for each expert.

Methods

Unlike previous studies on gaze biometrics, which rely on eye tracking data from artificial visual stimuli (e.g., jumping point of light) on a general-purpose computer display, our experimental data are sourced from radiologists viewing and interpreting mammographic images on medical-grade dual-head displays in clinical radiology reading room settings under typical lighting conditions. Study participant recruitment and data collection were done according to a protocol approved by the Oak Ridge Site-Wide Internal Review Board. All study participants signed an informed consent form.

Data Collection

Screening mammograms were selected from the Digital Database for Screening Mammography, a publicly available database (Heath et al., 1998). The selected cases were digitized using a LUMISYS scanner (50 microns per pixel, 12-bit grayscale) and included 50 malignant mass cases, 25 benign mass cases, and 25 normal cases. Each case included four images (two per breast).

A custom graphical user interface (GUI) software was developed to collect eye tracking data while viewing mammographic cases on dual-head five-megapixel Totoku medical-grade displays. The GUI included functions to zoom in and out, pan, and magnify each mammographic view on screen. Radiologists were able to change selective views of four images, the craniocaudal and mediolateral oblique views of both the left and the right breasts, or all four images on two full-screen displays.

The software was designed to control the eye tracking apparatus for streamlined synchronization of collected eye gaze data with the corresponding mammographic image. For eye tracking data collection, we utilized the H6 head-mounted eye tracking device from Applied Science Laboratories with a sampling rate of 60 Hz and an accuracy of within 0.5° of visual angle. Each image reader was instructed to make a thorough observation and decision on the displayed mammograms prior to marking findings. This process ensures that the marking process, via computer mouse clicks, did not distract the image readers' visual search and perceptual activities during the screen process.

The data collection process was conducted over multiple sessions depending on each individual image reader's availability and preferences. Prior to every reading session, a nine-point eye tracker calibration was performed to ensure proper eye gaze data collection. Readers were permitted to pause and resume the experiment at any time to avoid visual strain or cognitive fatigue. The number of sessions per reader and average time per reading are summarized in Table 1. Snapshot of the eye tracking data collection in radiology reading room and replay of eye gazes is shown in Figure 1.

Gaze velocity was calculated by a two-consecutive-point central difference between raw eye gaze points. We consider eye gaze data only when both pupil and corneal

Table 1
Number of pause and resume, and average time spent for reading a case by readers grouped by their expertise level: new radiology residents (NR), advanced radiology residents (AR), and expert radiologists (ER).

Reader	Number of Sessions	Average Reading Time Per Case (Seconds)
NR 1	2	32.42
NR 2	3	28.41
NR 3	3	42.92
AR 1	2	49.39
AR 2	3	27.13
AR 3	2	21.15
AR 4	2	58.59
ER 1	2	33.73
ER 2	5	38.31
ER 3	4	69.66

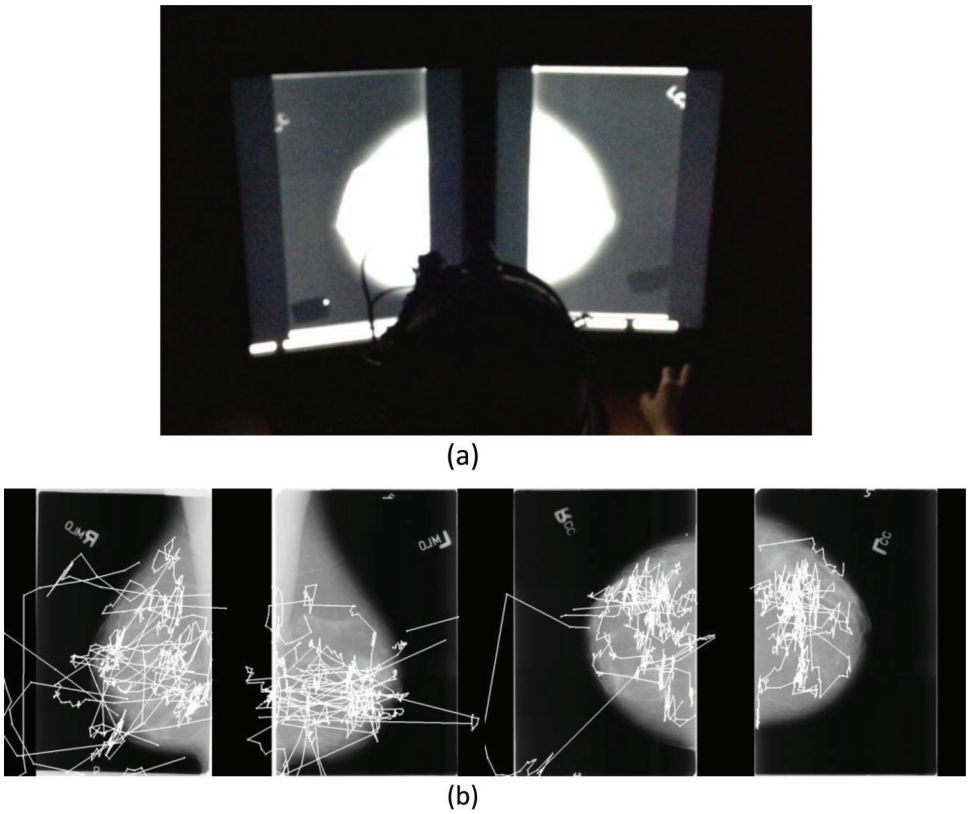

Figure 1. Eye tracking data collected by the custom graphical user interface (GUI) software and the head-mounted eye tracking apparatus. (a) Screening mammographic images read by 10 readers of radiologists in a dark radiology reading room. The GUI provided function to change selective views of four images, the craniocaudal (CC) and mediolateral oblique (MLO) views of both the left and the right breasts, or all four views on two medical-grade displays. (b) Example of eye gaze locations translated into image pixel coordinate and gaze scan path.

reflections were properly detected. Failures in detection of pupil and corneal reflections were mostly due to readers' blinking of his/her eyes; we do not calculate gaze velocities in between the failures. An illustration of gaze velocity for a three-second window is provided in Figure 2.

Classification of Temporal Sequences

There are several ways to extract characteristics from a temporal sequence. For eye tracking data, proposed approaches include frequency analysis using the fast Fourier transform (Kinnunen et al., 2010), statistical analysis of velocity (Silver & Biggs, 2006), and morphological analysis based on graph-based representations (Rigas, Economou, & Fotopoulos, 2012) to name a few. In our previous studies (Yoon et al., 2014, 2015), we employed HMMs, a probabilistic model with promising performance for classifying independent temporal sequences. HMMs capture temporal sequence dynamics using a state transition matrix, based on the Viterbi algorithm (Viterbi, 1967). Ten HMMs were trained (one HMM for each radiologist) and the log-likelihood from each HMM network was obtained given a test eye gaze velocity. The test velocity was assigned to the HMM with the highest log-likelihood.

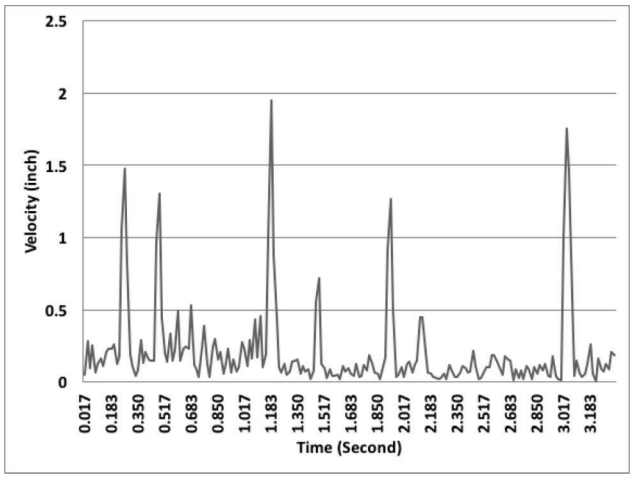

Figure 2. Example of eye gaze velocity of a radiologist reading a mammographic image on high-resolution displays. The eye tracking sampling rate is 60 Hz.

The optimal number of hidden states for the HMM network for this study was empirically determined based on study objectives. This approach is consistent with previous similar studies (Rothkopf & Pelz, 2004; Salvucci & Goldberg, 2000). Salvucci and Goldberg (2000) applied

two hidden states to represent fixations and saccades, and Rothkopf and Pelz (2004) proposed four hidden states to model smooth pursuit in addition to vestibular ocular reflexes. The number of optimal states was determined by monitoring the log-likelihood values of the HMMs. Using gaze velocity from the given training samples, we trained HMMs of two, three, four, and five hidden states for each radiologist, and selected the HMM with the highest log-likelihood. We applied an adaptation rule to find the optimal number of hidden states for each radiologist to avoid convergence failure and achieve better identification performance.

Deep Gaze Velocity Analysis

In this study, we propose a novel method for analyzing gaze velocity using deep learning algorithms. A deep learning algorithm is a type of machine learning algorithm, which uses multiple layers of nonlinear transformations to model high-level abstractions of large-scale complex data in an unsupervised fashion. This ability offers a unique advantage for eye tracking data analysis since, hitherto, feature engineering has relied primarily on expert domain knowledge.

Deep learning algorithms were trained to recognize associations between individual image readers and their characteristic gaze velocity profile. A typical deep learning network uses a softmax nonlinearity equation to compute the probability p_k for each output h_k where $N = 10$ radiologists. The output class label is assigned to the node with the highest output probability:

$$\max_k(p_k) = \max_k\left(\frac{\exp(h_k)}{\sum_{n=1}^{N}\exp(h_n)}\right) \quad (1)$$

Deep neural networks

A DNN is a feed-forward artificial neural network that has more than one hidden layers between its input and output layers (Hinton et al., 2012). DNNs are trained using a stochastic gradient descent (SGD) algorithm (Bottou, 1998), where node weights are estimated and updated using derivatives computed from small random subsets of training cases.

The large number of hidden layers in DNNs provides great flexibility, which makes them capable of modeling very complex and highly nonlinear data. This is a desirable property for modeling gaze velocity data. On the other hand, it also increases the possibility of modeling spurious patterns specific to the examples in the training set, which can lead to severe overfitting. Since the human subject data are expensive to collect, the DNN may have less chance to avoid the issue with limited number of training data.

Deep belief networks

DBNs (Hinton, Osindero, & Teh, 2006) stack multiple layers of restricted Boltzmann machines (RBMs), which are unsupervised nonlinear feature learners based on a probabilistic model. RBMs are unsupervised, probabilistic feature learners, capable of learning the right features in an unsupervised fashion (i.e. feature extraction and selection steps are not necessary). DBNs are trained using gradient approximation algorithms such as contrastive divergence approximation. This method treats the network input as an undirected Markov chain, which is sampled using Gibbs sampling.

DBNs are trained using a layer-based greedy algorithm. First learn all the weights for all layers combined, then freeze the bottom layer RBMs. The activations of the trained features are then treated as inputs for the remaining layers. The learning process is repeated on the remaining layers by iteratively learning and freezing the lowest (bottom) layer until only the top layer is left. This approach is guaranteed to improve the generative model (Salakhutdinov & Hinton, 2009). The output of the top hidden layer is expected to approximate the posterior for all the hidden units at all levels. Based on the output of the top hidden layer, the input sample is classified using a logistic regression classifier.

However, the DBN is referred to as a time-variant feature learner, because each RBM layer is fully connected to its input. Since the events of gaze fixations and saccadic movements are independent of time, the time-variance property of DBN makes it less desirable for the gaze velocity analysis.

Convolutional neural networks

CNNs are regarded as a variant of DNNs. They consist of an alternating convolution and pooling of layers followed by fully connected layers (Krizhevsky et al., 2012). Unlike traditional neural networks, where neurons are fully connected between layers, kernels (or filters) on a convolution layer have connections only at a local region in the input. This design causes each neuron within a convolution layer to focus on local properties representing time-invariant features. The output of neurons in a convolution layer is used as input to the next convolutional layer, allowing the network to detect more abstract, higher-level features. Neurons in the fully connected layers receive inputs corresponding to feature representations in the convolution layers. These fully connected layers are trained and used as the same manner as is prescribed in regular neural networks. The fully connected layers learn high-level abstractions.

There are no formal rules on how to choose the size of the convolution layer and the max pooling layers yet, but some recommended guidelines for designing CNNs do exist. For example, Szegedy, Vanhoucke, Ioffe, Shlens and Wojna (2016) suggest applying multiple stacks of smaller receptive field convolutional layers rather than one large receptive field convolutional layer, to achieve parameter efficiency and introduce more nonlinearity. In this study,

we propose a CNN topology inspired by the common formalization of fixation and saccadic analysis. Specifically, our CNN topology included the first convolution layer to capture gaze velocity characteristics during a fixation (1 × 6), where 6 samples is equivalent to the minimum duration of a fixation (3 consecutive eye gaze samples within an acceptable radius, and 3 subsequent eye gaze samples). Likewise, the second convolution layer was applied to capture clustered fixations (1 × 5). Three additional layers were added to reduce the dimensionality of the gaze velocity representation.

Evaluation

We calculated the F_1-score to evaluate the performance of the methods. The F_1-score is widely used to evaluate multiple-class classification in information retrieval. The F_1-score is the harmonic mean of precision and recall, which are computed as follows.

Let $TP_i = M_{ii}$, $FN_i = \sum_{j \neq i} M_{ji}$, and $FP_i = \sum_{j \neq i} M_{ij}$, where M_{ji} implies the number of decisions to reader j for given gaze velocity samples belonging to reader i. Then,

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \tag{2}$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{3}$$

The F-score for reader i can be calculated as:

$$F_\beta\text{-}score_i = \left(1 + \beta^2\right) \cdot \frac{Precision_i \cdot Recall_i}{\beta^2 \cdot Precision_i + Recall_i} \tag{4}$$

where β weights importance to recall if $\beta > 1$ and to precision otherwise. Typically, $\beta = 1$, called balanced *F-score*, or *F_1-score*

Macro-average F_1-score is defined as

$$Macro - F_1 = \frac{1}{N} \cdot \sum_i F_1 - score_i \tag{5}$$

where $N = 10$ is the number of readers.

Experimental Setup

We performed experiments to compare the classification performance of the three deep learning techniques and benchmarked them against an HMM-based classifier. Sequences of gaze velocity data were trimmed or padded to have equal length of 3,600. Classification performance was evaluated using 10-fold cross-validation. At each fold, gaze velocity data from 90 cases were collected as a training set, and it was scored based on classification performance on the 10 remaining test cases.

Optimal network topologies and learning parameters of the deep learning classifiers were determined empirically. The optimally configured DNN was 3,600 × 7,200 × 7,200 × 3,600 × 100 × 10 nodes. Training was performed using the SGD algorithm, where the learning rate was 0.05 for 25 epochs. The DBN classifier included 3,600 × 1,800 × 900 RBM layers. The learning rate for pre-training of the RBMs was 0.005 for 1,000 epochs, and the learning rate for fine-tuning was 0.005 for 1,000 epochs.

The CNN classifier included five convolutional layers and two fully connected layers. The first convolution layer included 20 kernels of size 1 × 6, the second convolutional layer filtered the output of the first convolution layer with 40 kernels of size 1 × 5, the third layer with 80 kernels of size 1 × 4, the fourth layer with 80 kernels of size 1 × 3, and the last convolutional layer with 80 kernels of size 1 × 2. The fully connected layers were 1,024 × 128 × 10. The SGD algorithm was applied with a learning rate of 0.08 for 250 epochs. Figure 3 illustrates the network topology of the three deep learning classifiers.

The deep learning classifiers were implemented with Theano (Al-Rfou et al., 2016) and Torch (Collobert, Kavukcuoglu, & Farabet, 2011) deep learning libraries. Experiments were executed at the Oak Ridge Leadership Computing Facility on the Titan supercomputer with NVIDIA Tesla K20 GPU accelerators.

For the implementation of HMM classifiers, ten HMMs were trained, one for each radiologist, and the log-likelihood from each HMM was computed for a given test gaze velocity. The test gaze velocity sequence was assigned to the radiologist whose HMM gave the highest log-likelihood output. The HMM classifier was implemented using the Pedregosa et al. (2011) package on Python 2.7.

Results

Table 2 lists the F_1-scores of individual radiologists grouped by their level of expertise, new radiology residents (NR), advanced radiology residents (AR), and expert radiologists (ER), with the various classification algorithms followed by the macro-averaged F_1-scores of all reader identification performances.

As shown in Table 2, the experiment confirmed the presence of personal differences during human visual perception and recognition activities. Still, the results clearly demonstrated that the CNN performed substantially better than the other algorithms. The competitive advantage demonstrated by CNNs across all individuals suggests that the local feature representation of one-dimensional kernels was a very effective way to capture individual characteristics reflected in eye gaze patterns. The HMM classifier-based method performed competitive to the CNN classifier for a few participants (i.e., NR2, ER2). On the contrary, the DNN and DBN classifiers performed poorly. It was observed during the DNN training that the training accuracy for a particular training set

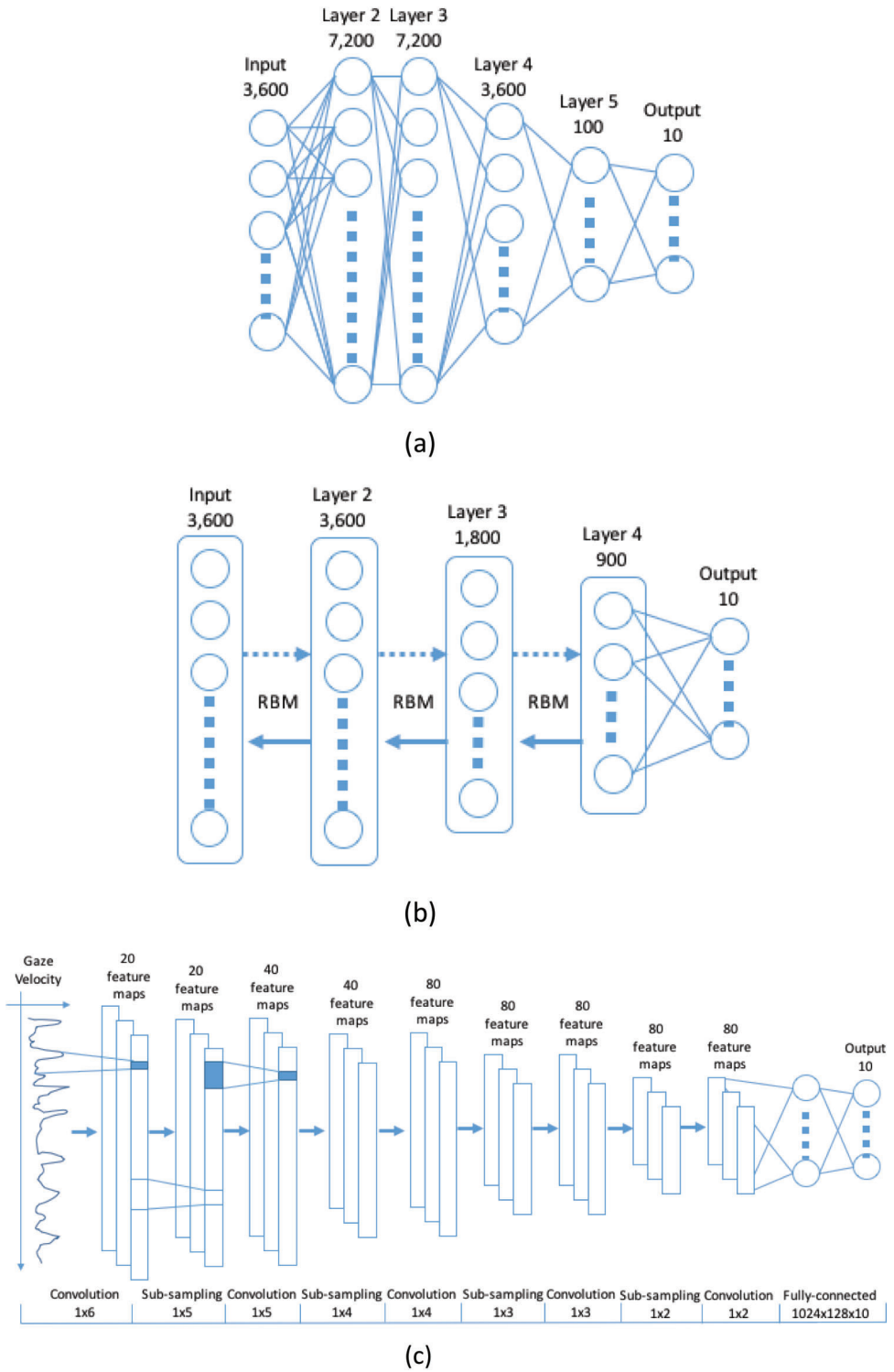

Figure 3. Neural network topology of the three deep learning methods applied in this study: (a) deep neural network, (b) deep belief network, and (c) convolutional neural network.

reached 100%; however, the accuracy to the corresponding testing set remained low. This finding can be regarded as evidence that the DNN classifier was overfitted to the training data. The DBN classifier performed the worst among all the methods failing to represent effectively the time-variant characteristics of the individual gaze velocity patterns.

The CNN classification accuracy was not only superior but also more robust across all radiologists compared to the HMM-based method. For example, the classification results of radiologists NR1, AR1, and ER3 were critically low with the HMM classifier, suggesting that the HMM completely failed to model the gaze velocity pattern for

Table 2
Measurement of classification performance in macro F_1-scores of the classification algorithms: HMM, DNN, DBN, and CNN. Note that the chance level is 0.1.

Readers	F_1-Scores			
	HMM	DNN	DBN	CNN
NR1	0.111	0.238	0.056	0.874
NR2	0.741	0.247	0.177	0.725
NR3	0.201	0.129	0.135	0.457
AR1	0.018	0.091	0.070	0.558
AR2	0.667	0.118	0.079	0.850
AR3	0.278	0.161	0.019	0.751
AR4	0.547	0.543	0.312	0.785
ER1	0.438	0.165	0.205	0.556
ER2	0.578	0.314	0.158	0.607
ER3	0.053	0.084	0.167	0.535
Macro-F_1	0.363	0.209	0.138	0.670

these radiologists. It is evident though that the local feature representation capacity of the CNN-based classification method was very effective for the same task.

Discussion

This paper explored the feasibility of eye gaze velocity as a behavioral, "soft" biometric. Eye tracking data were collected from radiologists reading digitized screening mammographic cases. Among the tested classifiers, the CNN-based classification method performed substantially better than the other three methods included in the analysis. Our study findings demonstrated that the feature representation capacity of the convolution layers which possess time-invariance was very effective when applied to gaze velocity sequences. Therefore, the CNN classifier is an effective means of analyzing one-dimensional temporal gaze sequence data.

Note that the study participants were trained experts performing a complex visual search task. The classification results confirmed that there are distinct personal differences in visual search for cancer detection tasks. However, even though the classification accuracy of the CNN-based classifier was noticeably higher than that of other competitive classification methods, it was still substantially variable across all radiologists, ranging from 0.457 to 0.874. Because data collection was done in well-controlled environments, it can be assumed that the variability may be induced by personal factors, such as vision condition, visual fatigue, cognitive burden, or duration of reading session. In future studies we will investigate the generalizability of our findings across various reading conditions to improve upon the robustness of eye gaze as a behavioral biometric.

In summary, humans often perform risk-sensitive decision tasks that require simultaneous visual processing and high-level cognitive integration of complex multimodal visual information such as images or videos from multiple monitors (e.g., air-traffic management and control in busy airports, time-critical military battlefield management, and diagnosis of life-threatening diseases from multimodality medical imaging data). To maximize human performance, intelligent user interfaces and decision support systems must be developed to account for the individual's perceptual and cognitive limitations to effectively synthesize rich visual content in a time-efficient manner. Although our study focused on a clinical visuo-cognitive task and medical experts, the general findings and developed approaches are easily extensible to other application domains. Our study is an important first step to characterize an individual's perceptual behavior and develop a data-driven "perception to cognition" framework for modeling human cognitive performance in complex visual environments.

Acknowledgments

The study was supported by the Laboratory Directed Research and Development (LDRD) program of Oak Ridge National Laboratory, under LDRD project No. 6268. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S., Department of Energy under Contract No. DE-AC05-00OR 22725.

References

Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., Ballas, N.,…Bengio, Y. (2016). Theano: A Python framework for fast computation of mathematical expressions. arXiv preprint arXiv: 1605.02688.

Bednarik, R., Kinnunen, T., Mihaila, A., & Fränti, P. (2005). Eye-movements as a biometric. *Image Analysis*, 16–26.

Bottou, L. (1998). Online learning and stochastic approximations. *On-Line Learning in Neural Networks*, *17*(9), 142.

Collobert, R., Kavukcuoglu, K., & Farabet, C. (2011). Torch7: A Matlab-like environment for machine learning. In *BigLearn, NIPS Workshop* (No. EPFL-CONF-192376).

Cuong, N. V., Dinh, V., & Ho, L. S. T. (2012, November). Mel-frequency cepstral coefficients for eye movement identification. In *2012 IEEE 24th International Conference on Tools with Artificial Intelligence (ICTAI)* (Vol. 1, pp. 253–260). Piscataway, NJ: IEEE.

Deravi, F., & Guness, S. P. (2011, January). Gaze trajectory as a biometric modality. In *Biosignals* (pp. 335–341).

Fookes, C., Maeder, A., Sridharan, S., & Mamic, G. (2009). Gaze based personal identification. *Behavioral Biometrics for Human Identification: Intelligent Applications*, 237–263.

Galdi, C., Nappi, M., Riccio, D., Cantoni, V., & Porta, M. (2013, June). A new gaze analysis based soft-biometric. In *Mexican Conference on Pattern Recognition* (pp. 136–144). Berlin, Germany: Springer.

Heath, M., Bowyer, K., Kopans, D., Kegelmeyer Jr, P., Moore, R., Chang, K., & Munishkumaran, S. (1998). Current status of the digital database for screening mammography. In *Digital mammography* (pp. 457–460). Dordrecht, Netherlands: Springer.

Henriksson, N. G., Pyykko, I., Schalen, L., & Wennmo, C. (1980). Velocity patterns of rapid eye movements. *Acta Oto-Laryngologica*, *89*(3–6), 504–512.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N.,…Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82–97.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554.

Holland, C. D., & Komogortsev, O. V. (2012, September). Biometric verification via complex eye movements: The effects of environment and stimulus. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (pp. 39–46). Piscataway, NJ: IEEE.

Kasprowski, P., & Ober, J. (2004, May). Eye movements in biometrics. In *International Workshop on Biometric Authentication* (pp. 248–258). Berlin, Germany: Springer.

Kinnunen, T., Sedlak, F., & Bednarik, R. (2010, March). Towards task-independent person authentication using eye movement signals. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 187–190). New York, NY: ACM.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Proceedings of 25th International Conference on Neural Information Processing Systems* (pp. 1097–1105). New York, NY: ACM

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Maltoni, D., & Jain, A. (Eds.). (2004). Biometric authentication. *ECCV 2004 International Workshop, BioAW 2004* (Vol. 3087). New York, NY: Springer Science & Business Media.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O.,…Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rigas, I., Economou, G., & Fotopoulos, S. (2012). Biometric identification based on the eye movements and graph matching techniques. *Pattern Recognition Letters*, *33*(6), 786–792.

Rigas, I., Komogortsev, O., & Shadmehr, R. (2016). Biometric recognition via eye movements: Saccadic vigor and acceleration cues. *ACM Transactions on Applied Perception*, *13*(2), 6.

Rothkopf, C. A., & Pelz, J. B. (2004, March). Head movement estimation for wearable eye tracker. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications* (pp. 123–130). New York, NY: ACM.

Salakhutdinov, R., & Hinton, G. (2009, April). Deep Boltzmann machines. In *Proceedings of 12th International Conference on Artificial Intelligence and Statistics* (pp. 448–455).

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (pp. 71–78). New York, NY: ACM.

Schmidt, D., Abel, L. A., DellOsso, L. F., & Daroff, R. B. (1979). Saccadic velocity characteristics: Intrinsic variability and fatigue. *Aviation, Space, and Environmental Medicine*, *50*(4), 393–395.

Silver, D. L., & Biggs, A. (2006). Keystroke and eye-tracking biometrics for user identification. In *Proceedings of the 2006 International Conference on Artificial Intelligence* (pp. 344–348).

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818–2826). Piscataway, NJ: IEEE.

Tourassi, G. D. (2005). Current status of computerized decision support systems in mammography. In *Intelligent paradigms for healthcare enterprises* (pp. 173–208). Berlin, Germany: Springer.

Tourassi, G., Voisin, S., Paquit, V., & Krupinski, E. (2013). Investigating the link between radiologists' gaze, diagnostic decision, and image content. *Journal of the American Medical Informatics Association*, *20*(6), 1067–1075.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, *13*(2), 260–269.

Voisin, S., Pinto, F., Morin-Ducote, G., Hudson, K. B., & Tourassi, G. D. (2013a). Predicting diagnostic error in radiology via eye-tracking and image analytics: Preliminary investigation in mammography. *Medical Physics*, *40*(10), 101906.

Voisin, S., Yoon, H. J., Tourassi, G., Morin-Ducote, G., & Hudson, K. (2013b, May). Personalized modeling of human gaze: Exploratory investigation on mammogram readings. In *Biomedical Sciences and Engineering Conference (BSEC)* (pp. 1–4). Piscataway, NJ: IEEE.

Yoon, H. J., Carmichael, T. R., & Tourassi, G. (2014). Gaze as a biometric. *Proceedings of SPIE*, *9037*, 903707.

Yoon, H. J., Carmichael, T. R., & Tourassi, G. (2015, March). Temporal stability of visual search-driven biometrics. *Proceedings of SPIE*, *9416*, 94160U.

Zhang, Y., & Juhola, M. (2012). On biometric verification of a user by means of eye movement data mining. In *Proceedings of the 2nd International Conference on Advances in Information Mining and Management (IMMM)* (pp. 85–90). Venice, Italy.