2013

University of North Carolina Wilmington

Master of Science in

Computer Science and Information Systems

Proceedings

https://csbapp.uncw.edu/mscsis

AN EVALUATION OF CLASSIFICATION PERFORMANCE FOR FACIAL ANALYTICS ON DIVERSE DATASETS

Jeffrey Nathen Raynor

A Thesis Submitted to the
University of North Carolina Wilmington in Partial Fulfillment
of the Requirements for the Degree of
Master of Science

Department of Computer Science
Department of Information Systems & Operations Management

University of North Carolina Wilmington

2013

Approved by

Advisory Committee

Laurie Patterson

Bryan Reinicke

Karl Ricanek, Jr.

Chair

Accepted by

Dean, Graduate School

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

ANN. Artificial Neural Network

ELM. Extreme Learning Machine

HOG. Histogram of Oriented Gradients

LBP. Local Binary Pattern

LOOCV. Leave-One-Out Cross Validation

SVM. Support Vector Machine

Abstract

An Evaluation of Classification Performance for Facial Analytics on Diverse Datasets. Jeffrey Nathen Raynor, 2013. Thesis Paper, University of North Carolina Wilmington.

Facial analytics is a derivative of face-based soft-biometrics that allows a machine to generate information about a person automatically, extracting this information from their face. Facial analytic applications generate descriptive information about an individual's face, but does not attempt to identify the subject.The attributes generated from these systems may include face shape, face pose, age, sex and other identifiable information. This work examined different types of texture descriptors, to determine the best descriptor for determination of sex (gender) and race. Further, this work will compared three general purpose machine learning classification techniques. The feature descriptors examined were local binary patterns (LBP), histogram of oriented gradients (HOG) and GIST. Each feature descriptor was married to all of the following classifiers: Artificial Neural Networks, Support Vector Machines and Extreme Learning Machines to determine which combination of feature and classifier generates the best recognition performance for gender and race recognition on two well-researched face databases and one extremely large scale database that contains 1.2 million faces.

Acknowledgments

I would like to acknowledge the following, in no particular order, for their contributions to this work.

Dr. Karl Ricanek Jr.

Dr. Ricanek you have been one of the most influential professors that I have had the pleasure of encountering. You are one of the main reasons that I pursued graduate school and am greatly responsible for my success.

Dr. Laurie Patterson

You solely stopped me from making the biggest mistake of my life and I thank you that. You are one of the most caring professors that know and it has been a pleasure to work with you these past couple years.

Benjamin Barbour

Ben, you are one of the most driven people that I have ever met. Your ability to learn new technologies with ease is inspiring. Thank you for all your help and guidance through graduate school and this thesis work.

David Macurak and Michael Sodomsky

You both have been a huge help through my undergrad and graduate studies. Thanks for putting up with my stubborn ways and my sometimes overbearing personality.

Bruce, Debbie, Josh, and Jennifer Raynor

You all have been so inspiring sense day one. Thank you for all the love and support, this work is dedicated to you.

Chapter 1: Introduction

Humans have an innate ability to determine the gender and race of one another with high accuracy and at a remarkable rate; however this process is extremely tough to automate for machines. In order to understand how people are able to achieve such high performance, we must first identify the particular cues and methods that humans use. The human face conveys a wealth of information about a person. Paul Ekman mentioned "faces are accessible windows into the mechanisms which govern our emotional and social lives" [15]. With a quick glance, an individual can discern a number of attributes from a person such as mood, age, gender, and race [6, 45, 9]. These attributes are primarily driven by the craniofacial region (the combination of the skull and face region), which typically consists of wrinkles, creases, varying skin pigmentation, and varying geometric shapes [43]. However, these are not the only attributes that humans use to determine one's gender, age or ethnicity; additional cues are inferred. These cues may include attire, physique, or behavior. A computer does not possess the ability to benefit from these additional attributes. Humans are believed to possess the ability to gather more discriminating information from a face than originally believed. According to Bruce and Young (1986), people are able to derive seven distinct types of information from a face; "pictorial, structural, visually derived semantic, identify-specific semantic, name, expression and facial speech codes" [9]. It is believed, these codes, collectively, hold the functional relationship to recognizing faces.

After understanding the cognitive process a human performs to determine a classification (i.e. race, gender, etc.), it is apparent why classification problems for a machine are so complex. In order for a machine to complete a classification problem (i.e. gender classification), there must be a number of subsequent steps: image acquisition, face detection, face alignment, feature extraction, feature selection, and identification or classification [10].

To a computer, an image is inherently a three-dimensional matrix of pixels, where each pixel possesses a red, green and blue descrete value, and in some image formats, an

alpha value. In order for a computer to detect a face in an image, it must manipulate the pixels in an useful manner. Rather than using the RGB values from each individual pixel, which can be computational costly, many researchers use statistical representations from simple features [52, 53, 28]. According to Viola and Jones, "The most common reason is that features can act to encode ad-hoc domain knowledge that is difficult to learn using a finite quantity of training data" [52]. Viola and Jones developed a face detection algorithm that is currently implemented in the OpenCV (Open Source Computer Vision) [7] library, and used by many researchers [52, 53, 28]. Their face detection algorithm is based on Haar-like features. See figure 1.1 for a depiction of Haar features.

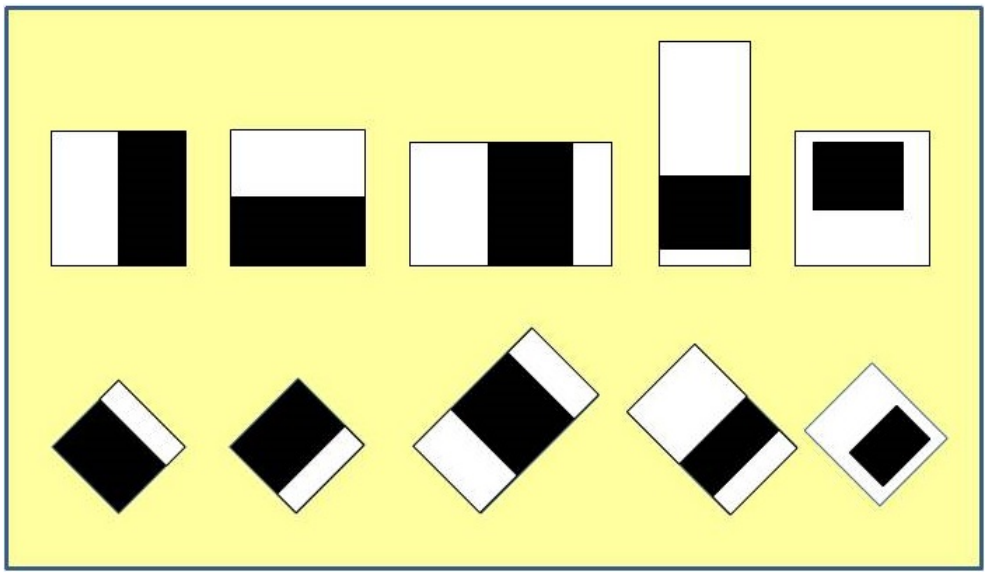

Figure 1.1: Haar-like Features

1.1 Previous Work

Golomb, Lawrence, and Sejnowski (1991) used neural networks to classify gender [17] . The experiment was tested by comparing eight tests of the network, each test trained on a different set of 80 faces, while leaving a different set of ten untrained faces for testing. The best results from the artificial neural network were recorded as error rates of 15, 0, 20, 0, 20, 10, 0 and 0 %, for an average error rate of 8.1%. However this was on a dataset that consisted of only 90 photos (45 male and 45 female).

Xu, Li, and Shi used a hybrid approach for gender classification [55]. Their method

includes fusing appearance and geometry features of the face. To extract the appearance features they used Haar wavelets and applied Adaptive Boosting (AdaBoost) to select the strong classifiers. To extract the geometry features, Active Appearance Models (AAM) were used. AAM generated 83 points on the face and extracted a total of 3,403 geometric texture features. The dataset used contained 14,756 images, with 8,544 male and 6,212 female. These images were taken from FERET [42] and AR [33] databases. The dataset was divided into 5 subsets, where each subset contained 800 images for training and 200 images for testing. Each training and testing subset had a uniform distribution of males and females, while each test set was independent of the training set. The best performance was achieved by classification using a Support Vector Machine (SVM) with the radial basis function (RBF) kernel. Their hybrid approach achieved a classification rate of 92.38% (error rate: 7.62%).

Mehmood, Ishtiaq, Tariq and Jaffar used ensemble based classifiers to classify gender [34]. They combined five different classifiers by using a Genetic Algorithm to search for the best solution. They tested with the Stanford University Medical Student (SUMS) dataset, which contained 200 male and 200 female gray-scale images with different facial expressions and races. Their optimized classifier was tested against many different algorithms: K-nearest neighbor (KNN), Artificial Neural Networks using back propagation (BPNN), Support Vector Machine (SVM), weighted majority voting, Linear Discriminant Analysis (LDA) and Fisher Linear Discriminate (FLDA). The best classifier achieved an classification rate of 94% (error rate: 6%), while SVM able to achieve an accuracy of 92% (error rate: 8%).

Ozbudak, Kirc, Cakir, and Gunes used Principle Component Analysis (PCA) and Fisher Linear Discriminants (FLD) to classify gender [40]. Their approach was to determine which facial features were best for gender classification. They experimented by masking portions of the images, such as the forehead, eyes, nose, lip, or chin, and then test using FLD. The dataset consisted of 480 face images, 240 male and 240 female, from the Stanford and FERET [42] databases. They were able to determine that the most usefully

feature for gender classification is the nose, which when masked, produced a 54.25% classification rate (error rate: 45.75%). The second best feature for gender classification was the forehead with an 64.5% classification rate (error rate: 35.50%).

Nazir, Ishtiaq, Batool, Jaffar and Mirza tackle gender classification using the Stanford University Medical Student (SUMS) database [38]. They first used Viola Jones face detector to get the bounding box of the face, used histogram equalization to normalize the image, and resized the face to a 32x32 pixel image [52]. Discrete Cosine Transform (DCT) was implemented for feature extraction, and the coefficients were fed into their K-nearest neighbor (KNN) classifier. They were able to achieve an accuracy rate of 99.30% (error rate: 0.7%); however, the demographics of the dataset is unknown and cannot be located.

Gender classification was also completed in by Mozaffari, Behravan and Akbari [37]. They combined appearance-based and geometric-based face features to classify gender. To extract the features Discrete Cosine Transform (DCT), Local Binary Pattern (LBP) and Geometric Distance Features (GDF) was implemented. Two databases were used for testing, the AR database [33], and a custom ethnic database developed by the authors. The AR dataset was composed of a total of 126 people, where 56 were female and 70 male. The ethnic dataset was built with the same distribution as the AR dataset but of Iranian people. They were able to achieve gender classification rates of 84.6% (error rate: 15.4%), 80.3% (error rate: 19.7%), for the ethnic and AR databases respectively, when LBP and DCT were fused together. Once all features were combined, they were able to achieve a classification rate of 97.1% (error rate: 2.9%) and classification rate of 96.0% (error rate: 4.0%) for the ethnic and AR datasets respectively.

Guodong and Guowang used a linear SVM to classify ethnicity [19]. The images were a subset from the MORPH-II dataset, using 21,600 images with equal representation of Caucasians and African Americans. They were able to achieve a 99.3% classification rate (error rate: 0.7%) with biologically-inspired features with the fusion of Orthogonal Locality Preserving Projections. They also tested on the entire dataset of MORPH Album 2 (55,000 images) for race classification for Black, White, Hispanic, Asian and Indian

ethnicities. Guodong and Guowang reported accuracies of 98.3%, 97.1%, 74.2%, 59.5% and 6.90% for Black, White, Hispanic, Asian, and Indian respectively.

In Roomi, Virasundarii, Selvamegala, Jeevanandham, and Hariharasudhan implemented race classification based off three facial features, specifically skin color, forehead area and lip color [46]. They combined the FERET [42] and YALE [2] databases for this experiment, and classified race into three groups; Caucasoid, Mongoloid and Negroid. The full dataset consisted of approximately 250 samples of size 45x45. The authors detected the dominate color of the face and projected the image to the YCbCr color space and compared the projection to a skin classifier. The forehead area and lip color were then extracted by using Sobel Edge detection. The dataset was then randomly divided for testing and training purposes. They achieved a classification rate of 81%.

In Lu, Jain, and others used Linear Discriminant Analysis (LDA) to classify race of either Asian or non-Asian [31]. The dataset used was composed from Yale [2], AR [33], AsianPF01 [14], and NLPR databases. It contained a total of 2,630 face images of 263 subjects (10 images per subject). The dataset was randomly divided with two-thirds used for training and one-third for testing, for 20 experiments. The average recorded classification rate was 96.3% with an LDA ensemble.

Tariq, Yuxiao, and Huang used silhouetted face profiles to identify gender and ethnicity [51]. The dataset was generated from 3D face data collected by Hu et al. [22]. The dataset contained 441 images, with ages ranging from 18-30, and divided into four categories: Black, East and Southeast Asian, South Asian and White. For matching they used a method that was based on the shape context of the silhouette. They tested their performance by making a training dataset that had an equal ratio of male and females of each ethnicity. This training set contained a total of 128 images. The accuracy achieved was 71.20% for gender classification and 71.66% for ethnicity classification.

Lyle, Miller, Shrinivas, and Damon were able to improve gender and ethnicity classification by utilizing the periocular region of the face [32]. Experimenting on the Face Recognition Grand Challenge dataset (FRGC), they were able to obtain an average gender

classification accuracy of 93%, 90%, and 94% for the right periocular region, left periocular region, and the whole face respectively. The ethnicity experiments also yielded similar results, right periocular, left periocular, and the face achieved 90%, 91%, and 92% respectively. The primary focus of this experiment was to show that the periocular region has certain discriminatory features that can be exploited for classification purposes.

Hosoi, Takikawa, and Kawade used a Gabor Wavelets Tranformation combined with retina sampling to extract features for the purpose of ethnicity estimation [21]. They collected a total of 1,991 face images, which was the fusion of personally collected images and the HOIP database [1]. The dataset contained three ethnicities, Asian, European, and African, consisting of 771, 774, 446 images respectively. The best recorded results were 96.3%, 93.1%, and 94.3% classification accuracies for Asian, European, and African, respectively.

Gender, age, and ethnicity classification was performed by Yang and Ai on three separate datasets: FERET [42], PIE [49], and a snapshot database [56]. The datasets contain 1,196 subjects (3,540 images), 68 subjects (696 images) and 9,000 Chinese snapshot images, for FERET [42], PIE [49], and the snapshot dataset respectively. The best recorded results for gender classification was using the images local binary pattern histogram (LBPH) with Real AdaBoost as the classifier with an accuracy of 93.3% and 91.1% against the untrained FERET [42] and PIE [49] datasets. Ethnicity and age also was classified the best when pairing the images LBPH and Real AdaBoost. They achieved accuracy measures of 93.2% only on the PIE dataset [49] for ethnicity and 92.12% and 87.5% for age on the FERET [42] and PIE [49] datasets, respectively. For the ethnicity experimenter, the snapshot and FERET [42] were combined for the training set. Each other experiment was completed with a 5-fold cross validation approach.

1.2 Objectives and Contributions

In this experiment, gender and race were classified, by pairing a feature extractor with a machine learning algorithm on a total of three datasets. The goal was to determine which feature extractor and machine learning algorithm performs the best, by a measure

of classification accuracy. To further the contribution, the datasets utilized are FGNET, MORPH Commercial and Pinellas County Sheriffs Office, and contained approximately 1,000, 156,000 and 1.2 million face images respectively. This is the first, and most comprehensive, evaluation of classification performance, on such a diverse collection of datasets.

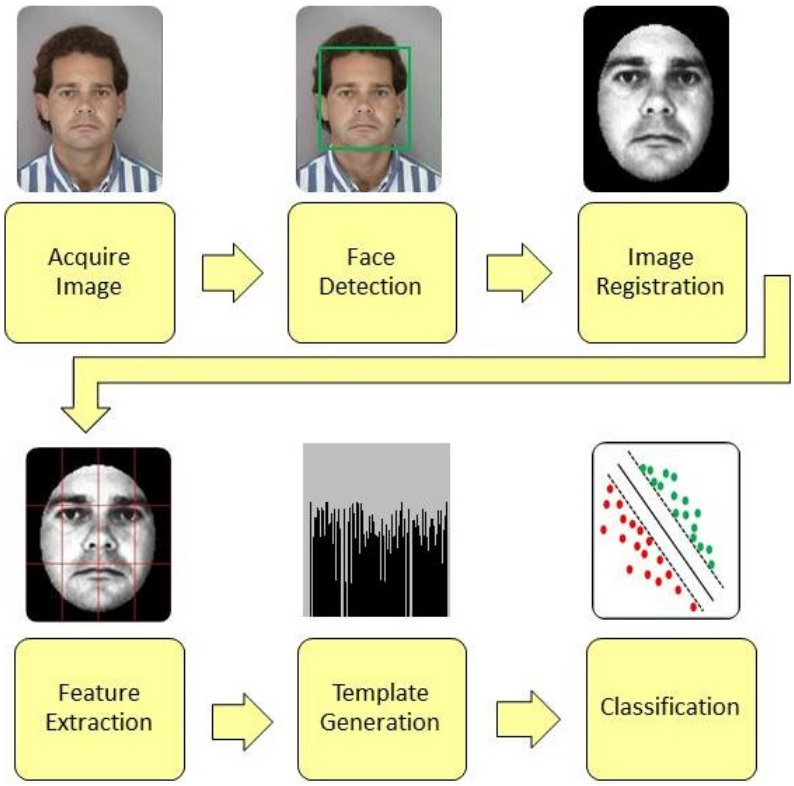

Figure 1.2: Workflow

This paper is structured as follows, Chapter 2 will give a general background of other researcher's work. Chapter 3 will explore each feature descriptors and machine learning algorithm that was implemented. My experimental approach with be described in Chapter 5. Chapter 6 will show the performance and results for each experiment, while Chapter 7 will summarize my findings and Chapter 8 will discuss room for future work.

Chapter 2: Background

2.1 Biometrics

Humans have used defining traits to identify others (such as face, voice, gait, etc.) since the beginning of time. This process was eventually given a name, biometrics. Biometrics is the ability to identify someone by their physical or behavioral characteristics. These characteristics may be their fingerprint, face, or even typing pattern. Possibly the first known use of biometrics was in China during the 14th century. Explorer Joao de Barros wrote about how the Chinese were using footprints to help distinguish their children from one another [50]. In 1858, Sir William Hershel, a civil servant from India, is documented for recording a hand-print on the back of each employees contract. This was used to ensure that each employee received their appropriate paycheck. Hershel's mechanism for distinguishing employees is known as "the first recorded systematic capture of hand and finger images that were uniformly taken for identification purposes" [30].

In 1870 Alphonse Bertillon, a Paris police chief, developed "Bertillonage" [50]. Bertillonage was a means to identify repeat criminals based on their geometric body measurements [27, 50]. It was not until 1903 that the Bertillon system was proven to be unreliable. A set of male, identical twins, were found to have the exact same Bertillon measurements. It was at this time that the Bertillon system was discredited and no longer implemented for person identification [50].

Anil Jain mentions that in order for a human characteristic to qualify as a biometric it must satisfy seven requirements: Universality, Distinctiveness, Permanence, Collectability, Performance, Acceptability, and Circumvention [27]. Universality is true when every person has the characteristic. This may be physical features such as the face or even behavioral features such as speech patterns. Distinctiveness applies when the characteristic is unique from one individual to another. Permanence is when the given characteristic does not change greatly over time. A good example of this would be the iris or fingerprint [1].

[1] Some research is investigating the permanence of the iris. Fingerprints can change in specific cases, e.g. manual laborers wear their fingerprints thin, which make them hard to acquire.

Collectability is the ability for the characteristic to be expressed as a form a measurement. Performance is the metric used to determine the effectiveness of the characteristic, for instance, recognition accuracy and speed. Acceptability is having the characteristic adopted as a biometric. Circumvention attributes to how easily the characteristic can be fraudulently reproduced.

2.2 Face Recognition Challenges

The most common challenges that exist for automated face recognition, are variations in illumination, pose, expression and occlusions [18, 4]. The representation, visually or statistically, can change dramatically due to these factors. Illumination can be defined as the intensity of light that is cast upon an object. In face recognition, as the amount of illumination varies, the descriptive information extracted from the face will deviate as well. Shadows that cast across the face can mask the face, which can lead to poor match performance or classification rates. However, psychophysical experiments have shown that humans can identify faces from vast changes in illumination [36], even with varying viewpoints [41, 13, 8, 36]. See Figure 2.3 for some example images with poor illumination.

Figure 2.3: Examples of poor illumination

To counter illumination variation, some researchers have attempted to model a face at different levels of illumination [4]. One image processing technique often used to normalize lighting is histogram equalization. For example, given a gray scale image x and n_i as the representation for the number of instances of the gray level i. The likelihood of an instance of the gray level i in x can be shown in Equation 2.1 Where L represents the

maximum number of gray levels in x, n corresponds to the total number of pixels in x, while $p_x(i)$ is x's normalized histogram from [0,1].

$$p_x(i) = p(x = i) = \frac{n_i}{n}, \quad 0 \leq i < L \tag{2.1}$$

The pose of a face can be defined as the manner at which the face is oriented relative to the body. The datasets used in this experiment consist of frontal-face images, meaning the subject is facing frontward toward the camera, both eyes visible, and the left hemisphere of the face shows the same amount of surface area as the right (see Figure 2.4 for examples of poor pose and Figure 2.5 for an example of a frontal-face image).

Figure 2.4: Examples of poor pose

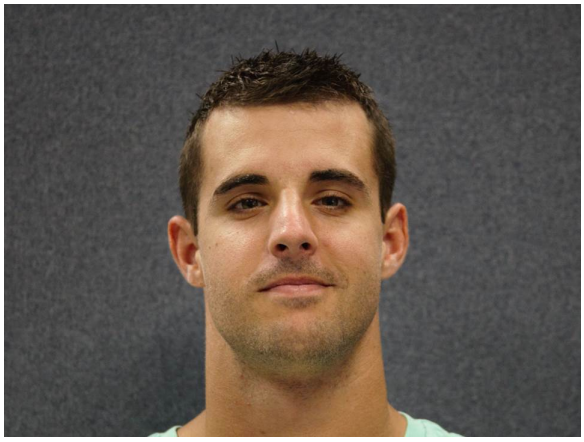

Figure 2.5: Example of a frontal-face image

Expression are the emotions and/or positioning of the face. Many expressions can

change the geometric shape of a person's face; happiness, for example, changes the face through smiling. As the geometric attributes of the face change, the feature descriptor's template will reflect the change as well. Due to this challenge, datasets often used for facial analytic experiments primarily consist of neutral expression images. See Figure 2.6 for examples of variating expression.

Figure 2.6: Examples of varying expression

A face can be masked by many things such as headgear, glasses, and even hair. When there is a barrier between the face and the sensor, the resulting signal is degraded relative to the same face without any occlusions. See Figure 2.7 for examples of occlusions.

Figure 2.7: Examples of occlusions

Soft-biometrics are techniques that use the biometric raw signal or the template signal to derive attributes that alone can not specifically identify a person. In the case of facial analytics, the biometric modality is the face and the biometric raw signal is the digital image of the face.

2.3 What Are Features

Features are an object's salient characteristics by which you can extract and exploit for certain tasks [26]. More specifically features from a human face are the attributes that compose the face, and can include wrinkles, creases, varying skin pigmentation, and varying geometric shapes [43]. Once certain features have been defined, a method for extracting them must be developed, which is referred to as feature extraction. According to Jain and Ricanek , "Feature extraction...involves the derivation of salient features from the raw input data in order to reduce the amount of data used for classification and simultaneously provide enhanced discriminatory power [26, 43]." There are many different algorithms used to extract features, such as local binary patterns (LBP), local ternary patterns (LTP), local derivative pattern (LDP), GIST, histogram of oriented gradients (HOG), scale-invariant feature transform (SIFT) and many others. These algorithms can be labeled as either: holistic (appearance-based) or local (region-based) methods. A holistic approach may use the entire set of pixel intensities or the histogram from the image, while a local approach is more focused on regions, such as geometric features (i.e. eyes, nose or mouth) [29].

3.0.1 Image Databases

The FGNET (Face and Gesture Recognition Research Network) aging database [3] is an publicly available face database that many researchers use for evaluational purposes. This database consists of 1,002 images of approximately 82 subjects, with an age range from 0 to 69. The majority of the images are scanned, old photographs and exhibit noise that is consistent with aged photography as well as variation of illumination, pose and expression [11]. See Figure 3.8 for sample images.

Figure 3.8: Sample Images from FG-NET Aging Database.

The MORPH (Craniofacial Longitudinal Morphological Face Database) database [44], is the largest publicly accessible longitudinal face database. These facial images have been captured from various individuals across a period of time. This dataset is a mug shot corpus, and hence, has some uncontrolled acquisition elements. MORPH is the combination of two of those datasets, Album 1 and Album 2.

MORPH Album 1 consists of 515 face images of individuals collected between October 26, 1962 and April 7, 1998. The images were acquired by scanning photographs

Figure 3.9: Sample Images from MORPH Album 2 Database.

with a consumer-grade scanner and then converting them into an 8-bit gray-scale portable gray map (PGM) format [44]. MORPH Album 2 is comprised of 56,000 facial images with ages ranging from 16 to over 60 [44]. However, the commercial version of MORPH contains approximately 156,000 face images with ages ranging from 16 to 77. The average timespan between photo acquisition is 164 days. See Table 8.11 for race, age range, and number of images per subject statistics [44].

The Pinellas County face dataset contains a total of 1,298,673 face images acquired from the Pinellas County Sheriff's Office in Florida, and is in a mug shot format. The images consist of approximately five different races with ages ranging from 1 to over 81, and are subject to some variations in illumination, pose, and expression. As a larger dataset, Pinellas County is the superior dataset to be used for evaluating facial analytic experiments. See Table 8.12 for Pinellas County's race, age range, and number of images per subject statistics. See Figure 3.10 for sample images.

3.0.2 Methodology

The research area of face-based soft-biometrics (facial analytics) is very popular.

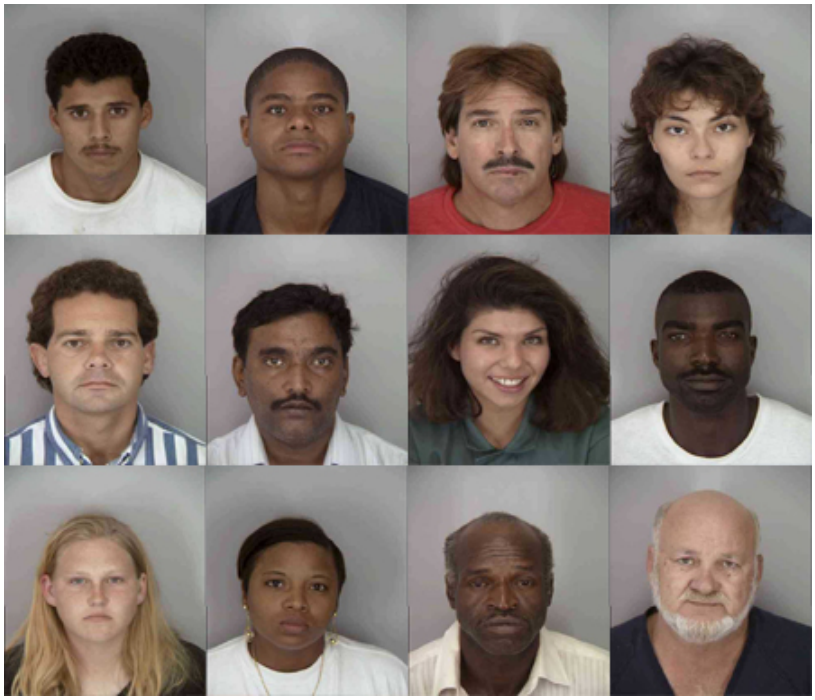

Figure 3.10: Sample Images from Pinellas County Database.

Many researchers have focused on finding the best feature encoding representation for the soft-biometric signal and fusing it with Support Vector Machines for the classification. In this work, I investigated two tried and tested feature encoding schemes: uniform LBP and histogram of oriented gradients (HOG), with a new technique for encoding features from scene analysis, GIST. These feature descriptors were applied to the face and paired with two general purpose classifiers: Artificial Neural Networks (ANN) and Support Vector Machines (SVM) as well with a novel general purpose machine learning technique, Extreme Learning Machines (ELM). It has been argued that ELM's are better general purpose classifiers. In this work, I investigated this claim as well as performed gender and race classification on two of the largest face corpora's available.

3.0.3 Face Pre-Processing

In order to generate the best possible results from gender and race classification, each image went through a number of pre-processing steps or registration. The registration process can be broken down into seven subsequent steps: monochromaticity, rotation, scale, translation, cropping, histogram equalization, and masking.

1. Monochromaticity

   In order to reduce the dimensionality of the classification problem each image was converted to grayscale. All images were converted to a grayscale representation, I', using the appropriate color transformation, see Equation 3.2.

$$I' = \frac{1}{3}(R, G, B) \tag{3.2}$$

2. Rotation

   The rotation step is the combination of a few steps prior to the actual rotation. First a face must be detected. This was accomplished by using OpenCV's [7] Viola Jones' face detector [52, 53, 28] as well as the face detector from Pittsburgh-Patterson Recognition SDK V5.2.2 [48]. Once a face is detected, the eye coordinates need to be documented. For images where a face was not detected by either face detector, a manual selection of the eye coordinates was performed. After obtaining the eye coordinates the image rotation can be completed. The rotation is completed by rotating the face image such that the eyes are horizontally aligned. See Equation 3.3 for for the rotation transformation matrix.

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3.3}$$

3. Scaling

   The image was then scaled accordingly, by fixing the interocular distance. This was accomplished through the use of the eye coordinates and a pre-determined interocular distance provided by the user (in this experiment the interocular distance was set to 40 pixels). This process was important as it ensured that each registered image had the same number of pixels between the eyes. Scaling, thus,normalized all face images.

The transformation matrix used for scaling can be found in Equation 3.4.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{3.4}$$

4. Translation

   The eye coordinates were also used during the translation process. After receiving the eye coordinates the translation offset was computed. This was a very important process as well, since it moved each eye location to the exact same position across all images. See Equation 3.5 for a sample translation matrix.

$$T_{\mathbf{v}} = \begin{bmatrix} 1 & 0 & 0 & v_x \\ 0 & 1 & 0 & v_y \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.5}$$

Figure 3.11: Artifacts from Translation

5. Cropping

Once all the former steps were complete, the image was cropped. This was completed by Equation 3.6, where r and c represents the number of rows and columns for new image I'. d represents the number of desired pixel between the eyes. The image was cropped to remove any artifacts that were present during the translation process (see Figure 3.11 for a visual representation).

$$I'(r,c) = \begin{cases} r = d * 1.5 + d \\ c = d * 3 \end{cases} \tag{3.6}$$

6. Histogram Equalization

   As mentioned earlier, illumination can be a problem with face recognition and classification problems. To mediate this challenge, histogram equalization was utilized, as discussed in detail in Chapter 2.2 and in Equation 2.1.

7. Masking

   After all the above steps were complete, the image was masked to focus on the region that was primarily concerned with (the face). The masking was automatically calculated using Equation 3.7, where r and c represent the number of rows and columns in the image. This process removed extra details that may be in the image such as headgear, jewelry, and facial hair.

$$Mask(x,y,width,height) = \begin{cases} x = c * 0.1 \\ y = r * 0.1 \\ width = c * 0.9 \\ height = r * 0.9 \end{cases} \tag{3.7}$$

The registration process was not perfect. If the face/eye detection returned an inaccurate location, the resulting registered image suffered accordingly. However, through visual inspection this process performed with high accuracy. See Figure 3.12, 3.13 and 3.14

for examples of the registration process for an image from FGNET, MORPH, and Pinellas County datasets, respectively.

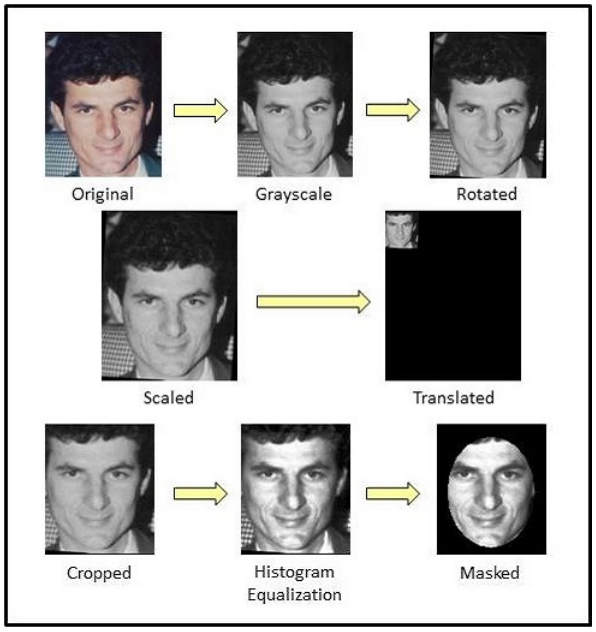

Figure 3.12: Example FGNET Image Registration

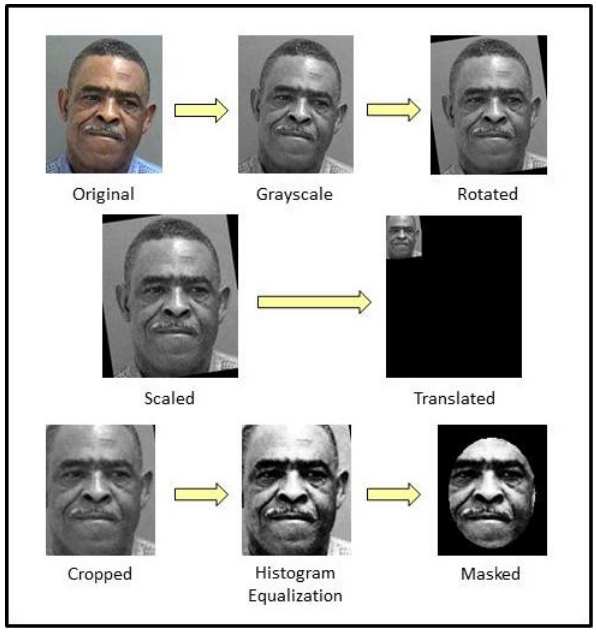

Figure 3.13: Example MORPH Image Registration

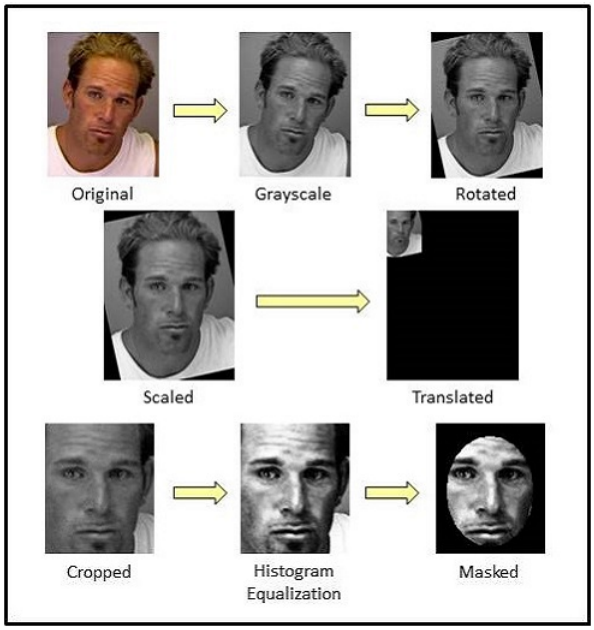

Figure 3.14: Example Pinellas Image Registration

3.0.4 Local Binary Patterns

Local Binary Patterns is a texture descriptor that can exploit a local or region-based approach to create discernible features [5]. This algorithm assigns a value to each pixel in a region by thresholding the sampled pixels against the center pixel of the region. Each neighborhood is then converted into a binary number and placed into the corresponding histogram bin. This process is completed from left to right and downward in a sweeping fashion until the final neighborhood is calculated. The generated histogram is then used as the texture descriptor of the image. A depiction of the LBP operator can be seen in Equation 3.8, where P and R represent the number of sampling points and the radius from the center pixel respectively. The LBP operator can work at different scales by adjusted the P and R values [5]. Equation 3.9 represents an LBP, with eight sampling pixels reaching only one pixel away from the center. Another variant of the LBP is called the uniform LBP. This operator performs the same as the original LBP but only looks for unique rotation-invariant patterns. The algorithm searches for these patterns by analyzing the number of bit shifts in the binary representation from a region and increases the count in the corresponding bin. Given that the uniform LBP searches for only unique patterns, the resulting histogram is

much smaller. More specifically with a (8,1) neighborhood, the uniform LBP generates a histogram with 59 bins compared to the original LBP's histogram bin of 256 (a 76.95% reduction) for a 8-bit image. In this work, the uniform LBP was implemented as one of the feature descriptors. See Figure 3.15 and 3.16 for an example of a LBP and uniform-LBP calculation respectively.

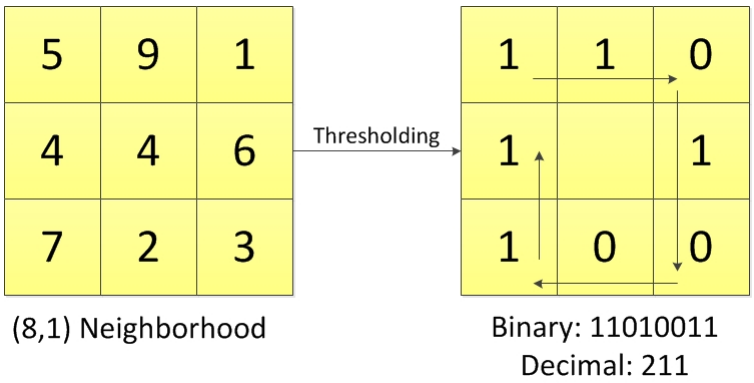

Figure 3.15: Example LBP Calculation

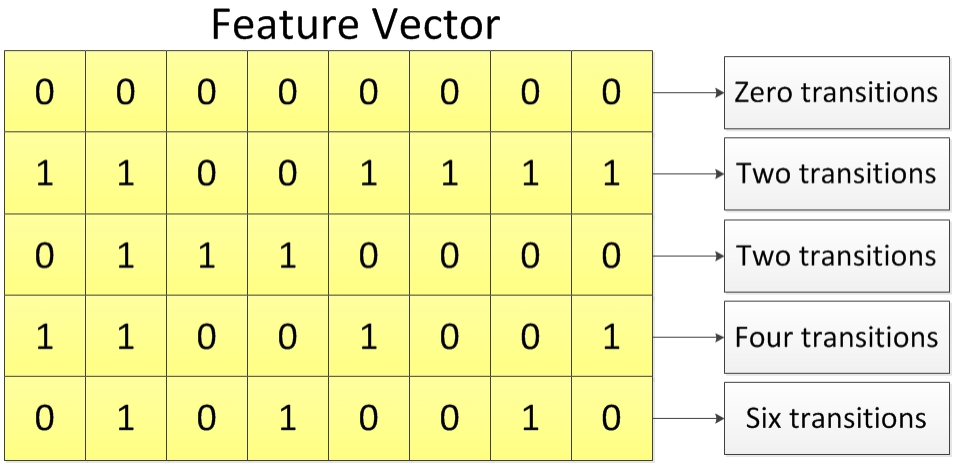

Figure 3.16: Uniform Patterns From Feature Vector

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \qquad s(x) = \begin{cases} 1, & if x >= 0; \\ 0, & otherwise. \end{cases} \tag{3.8}$$

$$LBP_{8,1} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \qquad s(x) = \begin{cases} 1, & if x >= 0; \\ 0, & otherwise. \end{cases} \tag{3.9}$$

3.0.5 Histogram of Oriented Gradients

Histogram of Oriented Gradient (HOG) was introduced by Navneed Dalal and Bill Triggs [12] as another local-based image descriptor. The premise of the algorithm is to use an image's distribution of edge directions as a means of object detection. According to Dalal and Triggs, the algorithm consists of a few subsequent steps: gamma/color normalization, gradient computation, spatial/orientation binning, and normalization and descriptor blocks. The end result of HOG returns a histogram of size 81 that can then be used for classification purposes.

3.0.6 GIST

GIST is another image descriptor but uses a holistic approach comparable to the other local approaches mention above. GIST is based on the idea that scene perception can recognized by a single glance [39]. This holistic algorithm is the interpretation of how the human visual system interprets a scene. More explicitly, "the visual system forms a spatial representation of the outside world that is rich enough to grasp the meaning of the scene, recognizing, a few objects and other salient information in the image, to facilitate object detection and the deployment of attention" [39]. For a more in-depth discussion on GIST consult [39] and [16].

3.1 Machine Learning

In 1959, Arthur Samuel defined machine learning "as the field of study that gives computers the ability to learn without being explicitly programmed" [47]. This definition was refined in 1998 by a Carnegie Mellon computer science professor, Tom Mitchell. Mitchell said "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

In this experiment, three machine learning algorithms (ANN, SVM, and ELM) were

chosen to be evaluated. The two main types of machine learning are defined as supervised and unsupervised learning. Both types were implemented in this experiment.

3.1.1 Supervised Learning

Supervised learning is an application of machine learning when the expected result or class is given during the training (learning) phase. In order to test the performance of the training, the expected output is compared with the trained algorithms output. Supervised learning can have multiple implementations such as classification or regression. Classification can be defined as the process of assigning a subject to a particular group or class. Classification can be a binary process (i.e. gender) or multi-class (i.e. age). However, regression attempts to locate a function that maps the inputs to the specific desired output. This technique can be used for determining the price of a particular stock in six months or even what the temperature will be at noon tomorrow.

3.1.2 Unsupervised Learning

Unsupervised learning attempts to construct an internal representation of the input data. During the training (learning) process, the desired output is not provided. This type of learning algorithm excels at finding hidden structure in data and can sometimes be referred to as clustering.

3.2 Artificial Neural Networks

The development of the neural network algorithm was inspired by how the brain is presumed to work. Scientist have adopted to use computer simulations to gather a better understanding of how the brain works. Even though we do not have a complete understanding of how the brain performs certain functions such as learning; in machine learning, the artificial neural network attempts to mimic the learning process of the brain.

The architecture of an Artificial Neural Network consists of layers, neurons, and weights. An Artificial Neural Network is comprised of three layers: input, hidden, and output. There can be only one input and one output layer while the number of hidden layers can vary. Each layer also contains a number of nodes. The nodes in the input layer directly correspond to the input set provided by some stimuli. The nodes at the output layer

represent the generated output from the network. Each node is also connected from layer to layer by a weight. This weight can be considered to be the connection strength from node to node. Each layer is also assigned an activation function that is used to calculate each nodes value. In order to do the calculation the layers activation function is applied to each connection from the previous layer and the value is then summed. This progresses until the output layer is reached and the output nodes are calculated. An example of a neural network can be seen in Figure 3.17. One the most common neural network paradigms is the feed-forward back-propagation algorithm. This particular algorithm accepts an input and pushes the values through the network to generate an output. The output is then compared to the expected output and the error is calculated. If the error is too high, the error is propagated backward through the network, adjusting the weights by the relative error. This is an iterative process that will continue to run until the calculated error is below a preset threshold or the maximum number of iterations has been achieved.

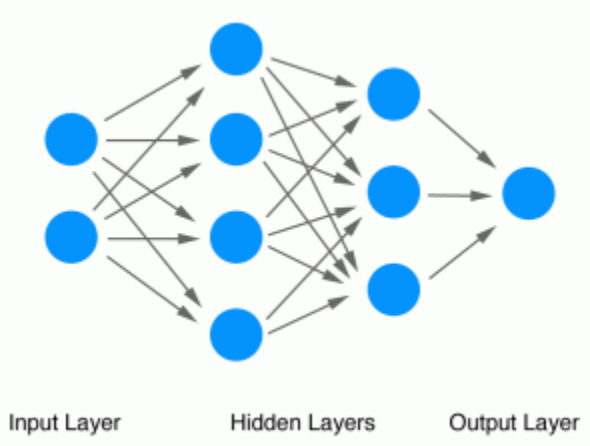

Figure 3.17: A 2-4-3-1 neural network

3.3 Support Vector Machines

The Support Vector Machine (SVM) algorithm uses a supervised learning approach that is used to detect patterns in data. The algorithm can be used as a means of classification or regression. It attempts to find a hyperplane or set of hyperplanes, in high or infinite-dimensionally space that will discriminate the data, from each class, by the highest margin. For a simple two-dimensional depiction of a trained SVM view Figure 3.18.

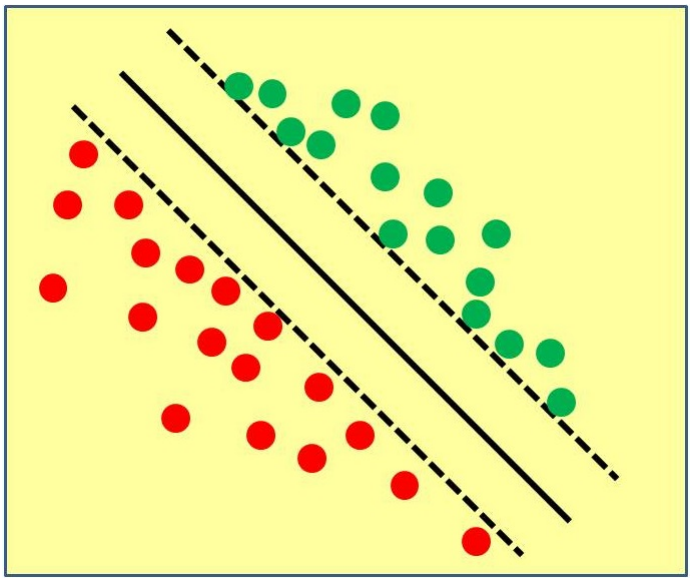

Figure 3.18: Support Vector Machine 2-D Hyperplane

3.4 Extreme Learning Machines

Extreme Learning Machines is a newer machine learning algorithm that is based on the single-hidden layer feedforward neural network (SLFN). A SLFN is an artificial neural network that contains only one hidden layer. The primary difference between SLFNs and ELMs is the fact that the hidden layer of the ELM does not need to be tuned [25, 24, 23]. Since the hidden layer isn't tuned, the learning process is much faster than that of the Artificial Neural Network (ANN). Just like the ANN, the ELM has the ability to be used as a means of classification or regression. In this experiment, ELM was used only for classification purposes, more specifically for gender and race classification.

4.1 Image Segmentation

As formerly discussed, the feature descriptors used in this experiment were both local and holistic approaches. The uniform-LBP and HOG descriptors were utilized in a different manner than the GIST descriptor. I segmented the image into three rows and four columns, yielding twelve subsections (see Figure 4.19 for a depiction of the image segmentation). The uniform-LBP and HOG were then calculated on each subregion from left to right and downward, beginning in the top left corner. Each subset's histogram was concatenated across each block yielding a total histogram bin size of 708 and 972 for the uniform-LBP and HOG respectively. Because GIST is a holistic approach, the image segmentation was not use during the feature extraction process, but did produce a histogram of bin size of 512. Provided are examples of the generated histograms for each feature descriptor from each dataset in Figure 4.20, 4.21, and 4.22.

Figure 4.19: Depiction of image segmentation

A better visual representation of the feature generation process (for uniform-LBP and HOG), appears in Figure 4.23. Each subsection has its corresponding histogram visually displayed.

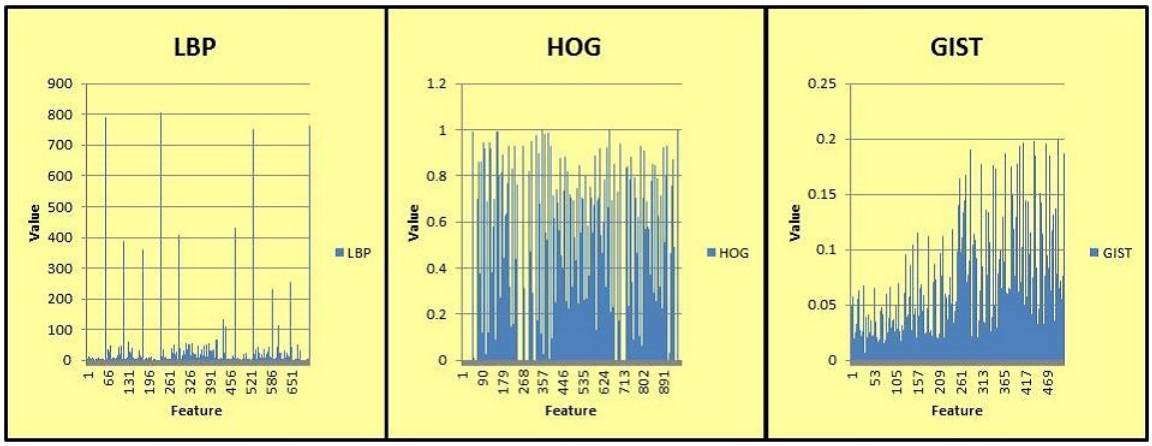

Figure 4.20: Example FGNET Histograms

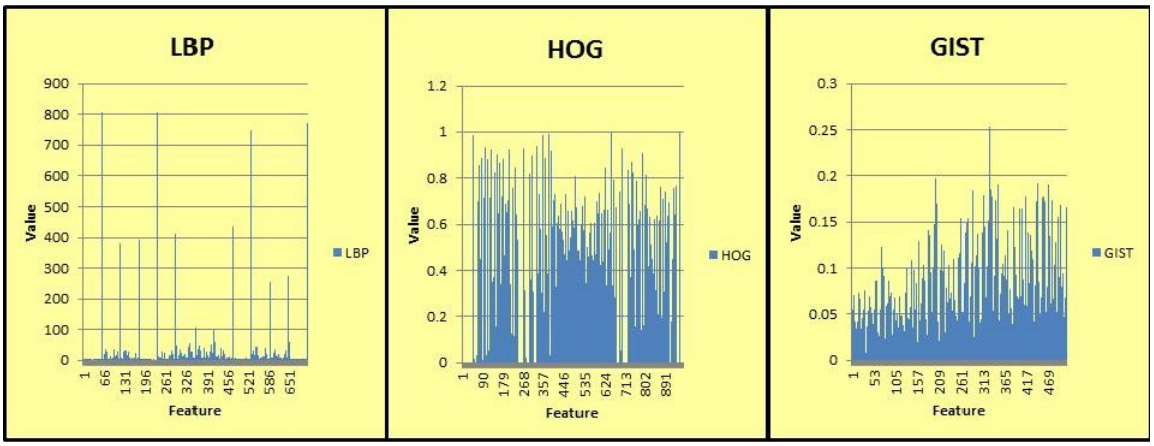

Figure 4.21: Example MORPH Histograms

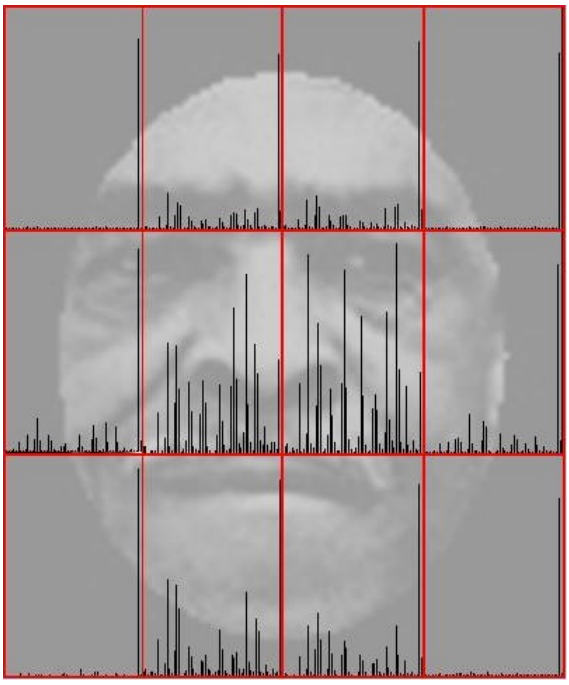

Figure 4.23: Histogram-ed Image

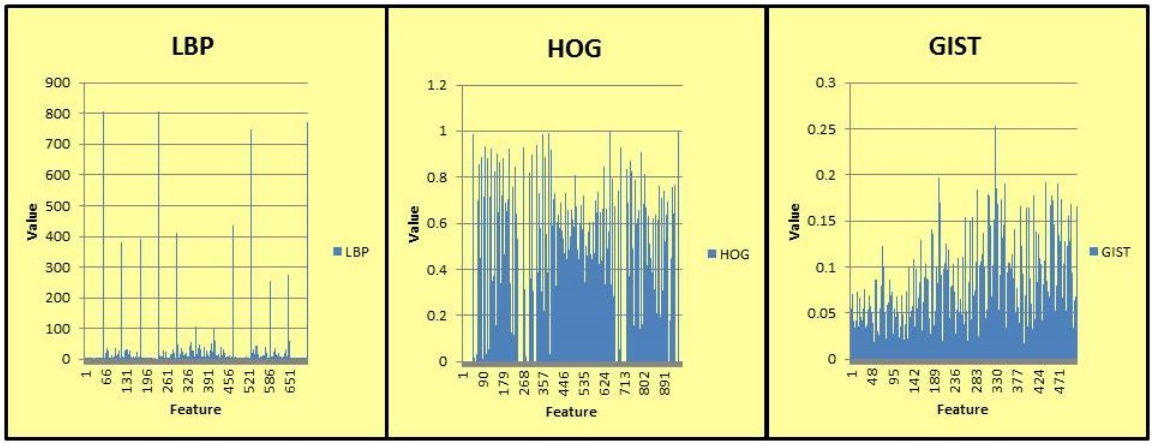

Figure 4.22: Example Pinellas Histograms

4.2 Dimensionality Reduction

Given the raw signals of size 708, 972, and 512, I needed to reduce the dimensionality in order to make the use of each machine learning algorithm computationally feasible. Principle Components Analysis (PCA) was used for dimensionality reduction. The registered face images used, to form the PCA projection function, were randomly chosen. The demographics for each can be found in Figures 8.30, 8.33, and 8.36, 8.31, 8.34, and 8.37 and 8.32, 8.35, and 8.38 for FGNET, MORPH, and Pinellas County respectively. Once the PCA was complete, the top 90% of the energy was retained.

4.2.1 FG-NET (Small Dataset)

Because the size of this dataset was small, the leave-one-out cross validation (LOOCV) was used (see Table 8.13 for a break-down of the size of each dataset). Implementing the LOOCV yielded 82 experiments (one for each unique subject). The training set contains all images of the 81 subjects and the testing set contains all images of the leave-out subject. This process is then completed again with a new unique subject and looped until all unique subjects have been used in the training set. An accuracy measure was generated for each subject\experiment and then combined for the overall accuracy. See Figure 4.24 for a visual depiction of leave-one -out cross validation with only ten unique subjects (labeled A through J).

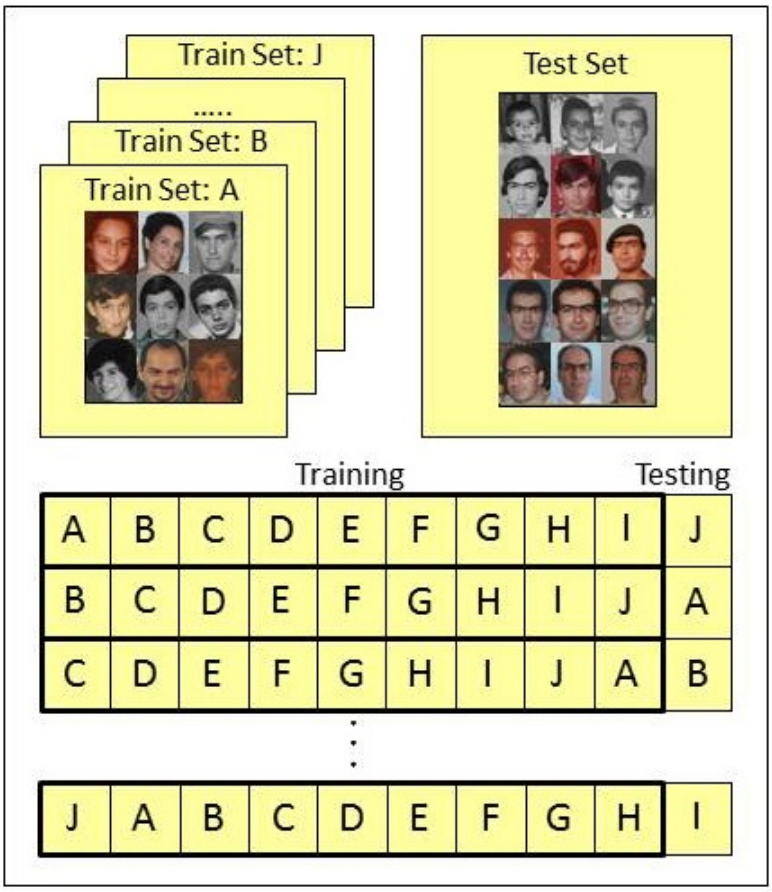

Figure 4.24: Leave-one-out Cross Validation

4.2.2 MORPH and Pinellas County (Large Datasets)

For the larger datasets, in order to generate the training and testing sets, I first derived the number of unique subjects and number of images for each corresponding subject across each datasets (see Table 8.13 for the total size of each dataset and the number of unique subjects per dataset). I then was able to sample accordingly where only unique subjects were in each training set with a uniform distribution (or as uniform as possible) between gender and race. Each training set was segmented into three parts: training, validation, and testing. For this experiment, 60% of the 'training' set was used for training, while 20% each was allocated for validation and testing. For example, the MORPH training set of size 1,266, Figure 4.25 visually displays the partitioning of the training, validation, and testing sets.

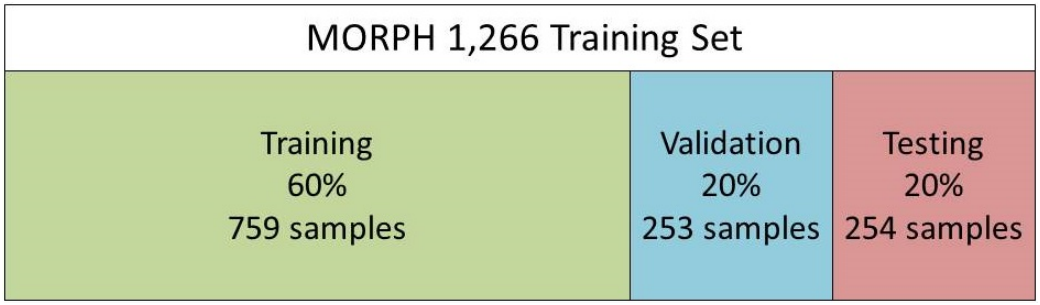

Figure 4.25: MORPH 1,266 Training/Validation/Testing Percentage

An evaluation set was also created to more fully test the generalization of each experiment. Provided in Figure 4.26 is the size of the evaluation set for the MORPH 1,266 dataset. The recorded accuracies in Chapter 6 and all appendices that are labeled as 'evaluation' correspond to the datasets evaluation set.

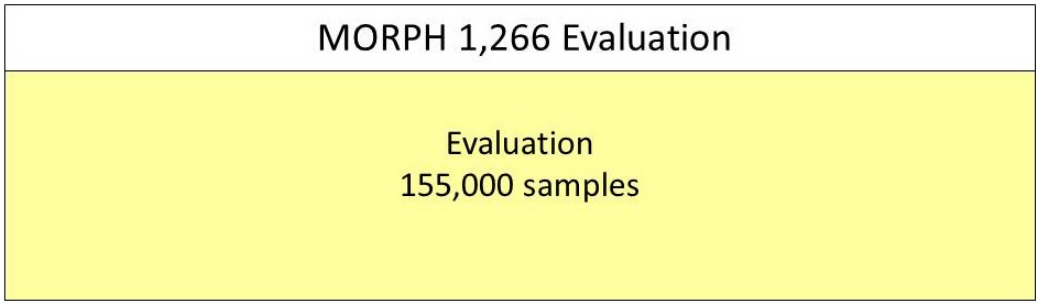

Figure 4.26: MORPH 1,266 Evaluation Set

A break-down of the demographics for each training\testing set can also be found in Appendix 2.

Great care was taken to ensure no bias was provided for any stage of this work, and hence, the pre-processing block, the small dataset versus the large dataset protocols, and the care placed into developing the training sets from sample the covariates (gender and race) are as equal a distribution as possible.

5.0.3 SVM Approach

The main two parameters to tune SVMs are the c (cost) and g (gamma) values. A gridsearch was implemented in search for the optimal parameters in relation to classification accuracy. In each SVM experiment, a gridsearch was completed with c and g values ranging from 2^{-25} - 2^{25} and 2^{-10} - 2^{10} respectively. The kernel used for all SVM experiments was the radial basis function (RBF) due to its popularity. The SVM experiments were implemented with only the training and testing set, while the validation set was unused. Once the training was complete, I selected the c and g values that yielded the best performance and tested the model against the evaluation set. Provided in Appendix 4 are the recorded accuracies for each test and evaluation set.

5.0.4 ANN Approach

Due to the vast number of parameters than can be tuned with ANN, only a few were adjusted to help find the most optimal network. To determine the topgology of the network, I chose a rule of thumb approach that consisted of three hidden layers, each layer's size resulting in half the size of the previous' layer (see table 5.1, 5.2, and 5.3 for each networks topology). The training function used was the Levenberg-Marquardt algorithm that was provided by MATLAB with the performance measurement of mean squared error. During the training portion of the ANN, the validation set was implemented to safeguard against the network over-fitting the data (this is noteworthy sense the SVM and ELM does not use a validation set). Each network was trained ten times to find the best network, from a classification perspective. I chose to run the network multiple times due to the random initialization of weights that could potentially not allow the network to jump out of any local minimas. From the ten trials, the best network was saved and then tested against

the evaluation set. Provided in Appendix 5 are the recorded accuracies for each test and evaluation set.

FGNET	**Network Topology**				
	Input Layer	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Output Layer
LBP	38	19	9	4	1
HOG	46	23	11	5	1
GIST	32	16	8	4	1

Table 5.1: FGNET Network Topology

MORPH	**Network Topology**				
	Input Layer	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Output Layer
LBP	98	49	24	12	1
HOG	167	83	41	20	1
GIST	63	31	15	7	1

Table 5.2: MORPH Network Topology

Pinellas	**Network Topology**				
	Input Layer	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Output Layer
LBP	106	53	26	13	1
HOG	167	83	41	20	1
GIST	60	30	15	7	1

Table 5.3: Pinellas Network Topology

5.0.5 ELM Approach

ELMs are claimed to be the 'out of the box', a no tuning required machine learning algorithm, given this claim little tuning was completed. The number of hidden nodes was set to 1,000 (suggested from the ELM developer, Huang Guang-Bin) and each activation function (Sigmoidal, Sine, Hardlim, Triangular Basis, and Radial Basis Function) was used

for each experiment. Like the ANN, I chose to run the ELM with ten trials to try to find the best performing ELM. Note, during the training process the validation set was not used, similarly to the SVM. After the ten trials only the best ELM was saved and then tested against the evaluation set. In Appendix 6 are the accuracies from each test and evaluation set are shown.

5.0.6 Cross Dataset Evaluation Approach

I also ran cross dataset evaluations for the MORPH and Pinellas datasets. I used only the trained models/networks from the MORPH 10446 and Pinellas 14322 datasets. For example, the MORPH 10466 gender model was used to evaluate the entire Pinellas dataset for gender classification. This was completed for each classifier (SVM, ANN, and ELM) with each feature descriptor (LBP, HOG, and GIST) for each experiment (gender, white, and black) for both datasets. In all fifty-four cross dataset experiments were completed; the accuracies can be found in Appendix 9.

To evaluate each pairing, I chose to compare the classification accuracies from each evaluation set. Due to the small size of FGNET, I omitted it from my final results in Tables 6.4, 6.5, and 6.6. The best ANN and ELM results for gender was 93.55% (Pinellas - HOG) and 92.60% (Pinellas - HOG) respectively. Specifically for the MORPH 1266 dataset, the ANN scored the highest accuracy of 92.07% (LBP), while the SVM performed the best on the MORPH 10446 set with 93.367%. The Pinellas dataset received the best accuracy from the HOG SVM with an classification accuracy of 94.81%, which also proved to be the best accuracy overall. The MORPH 1266 dataset's best accuracy was 92.82% from the HOG SVM for the white classifier. A 95.83% classification rate was the best performance on the MORPH 10446 dataset and best overall for white classification. The HOG SVM also performed the best on the Pinellas set with an 94.42% classification accuracy. The HOG SVM recorded the best results of 90.37%, 94.11%, and 96.12% for MORPH 1266, MORPH 10446, and Pinellas respectively for black classification. The overall best accuracy was 96.12% from the Pinellas dataset.

Gender Classification			
	MORPH 1266	MORPH 10446	Pinellas
SVM	87.87% (GIST)	93.37% (LBP)	**94.81% (HOG)**
ANN	92.07% (LBP)	82.19% (LBP)	93.55% (HOG)
ELM	85.36% (LBP - Tribas)	92.60% (LBP - Sig)	92.52% (HOG - Sin)

Table 6.4: Best Gender Classifications

White Classification			
	MORPH 1266	MORPH 10446	Pinellas
SVM	92.82% (HOG)	**95.83% (HOG)**	94.42% (HOG)
ANN	89.44% (GIST)	94.64% (HOG)	92.27% (HOG)
ELM	82.91% (LBP - Tribas)	93.91% (LBP - Sin)	91.63% (HOG - RBF)

Table 6.5: Best White Classifications

Black Classification			
	MORPH 1266	MORPH 10446	Pinellas
SVM	90.37% (HOG)	94.11% (LBP)	**96.12% (HOG)**
ANN	87.48% (HOG)	91.90% (LBP	94.91% (HOG)
ELM	79.11% (HOG - Sig)	92.66% (LBP - Sig)	95.61% (HOG - RBF)

Table 6.6: Best Black Classifications

The results for the cross dataset experiment are located appendix 9 and the best performing experiments can be found in tables 6.7 and 6.8. The HOG SVM trained models from the MORPH 10446 dataset performed the best on the Pinellas dataset across the board; boasting accuracies of 91.40%, 82.97%, and 96.13% for gender, white, and black respectively. When taking the trained models/networks from the Pinellas dataset and evaluating against MORPH, the best highest accuracy for gender was 93.58% with the HOG SVM. The LBP ANN performed the best for white classification with 86.51% accuracy and the GIST SVM recorded the best results for black classifications.

Best Performing Cross Dataset Classifiers		
MORPH Trained/Pinellas Tested		
	Classifier + Descriptor	Accuracy
Gender	SVM + HOG	91.40%
White	SVM + HOG	82.97%
Black	SVM + HOG	96.13%

Table 6.7: Best Performing Cross Dataset Classifiers (MORPH Trained/Pinellas Tested)

Best Performing Cross Dataset Classifiers		
Pinellas Trained/MORPH Tested		
	Classifier + Descriptor	Accuracy
Gender	SVM + HOG	93.58%
White	ANN + LBP	86.51%
Black	SVM + GIST	72.57%

Table 6.8: Best Performing Cross Dataset Classifiers (Pinellas Trained/MORPH Tested)

When analyzing all the classifiers and their paired feature descriptors, the best performing pairing overall was the HOG SVM. The better metric to gauge the performance of the pairing is analyze the cross dataset experiments. These experiments gave a more conclusive result of how well the model/network generalized to the data. Once again the HOG SVM stood out from the others and can be seen in Tables 6.7 and 6.8.

Chapter 7: Conclusion

This work examined different types of texture descriptors, determining the best descriptor for determination of sex (gender) and race. This work also compared three general purpose machine learning classification techniques: Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and Extreme Learning Machines (ELMs). The feature descriptors that were examined were local binary patterns (LBP), histogram of oriented gradients (HOG) and GIST. Each feature descriptor was paired with each classifiers; to determine which combination of feature and classifier generated the best recognition performance for gender and race recognition. The experiments were trained and evaluated on two well-researched face databases and one extremely large scale database that contains 1.2 million faces. Overall, the HOG SVM generalized the best for all three classification experiments. This pairing performed the best; however, I expect that with the addition of an intelligent feature selection the accuracies could be increased.

Chapter 8: Future Work

Facial analytics is an ever evolving field due to the development of newer feature descriptors and machine learning techniques. There are many different variants of the feature extractors that can be implemented to obtain a more conclusive derivative. Aside from implementing more feature descriptors, a closer examination of the image partitioning (with respect to feature extraction) could greatly further this work. From a visual inspect, the partitioning scheme chosen may not have been the best for this implementation. I believe there are certain regions of that face that yield more discriminative information than my current scheme was able to take full advantage of. Feature selection is a process that could increase the classification accuracies across each experiment. Applying a fusion of different feature descriptors is also a method that could boost the overall results.

References

[1] Hoip facial image database.

[2] Yale face database.

[3] Fg-net aging database, 2000.

[4] Y. Adini, Y. Moses, and S. Ullman. Face recognition: the problem of compensating for changes in illumination direction. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19(7):721–732, 1997.

[5] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 28(12):2037–2041, 2006.

[6] A. M. Albert, K. Ricanek, and E. Patterson. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. Forensic Science International, 172(1):1–9, 2007.

[7] G. Bradski. The opencv library. Doctor Dobbs Journal, 25(11):120–126, 2000.

[8] V. Bruce. Changing faces: Visual and non-visual coding processes in face recognition. British Journal of Psychology, 73(1):105–116, 1982.

[9] V. Bruce and A. Young. Understanding face recognition. British journal of psychology, 77(3):305–327, 1986.

[10] Y. Chang, Y. Wang, K. Ricanek, and C. Chen. Feature selection for improved automatic gender classification. In Computational Intelligence in Biometrics and Identity Management (CIBIM), 2011 IEEE Workshop on, pages 29 –35, april 2011.

[11] C. Chen, Y. Chang, K. Ricanek, and Y. Wang. Face age estimation using model selection. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pages 93–99, June.

[12] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 886–893. IEEE, 2005.

[13] G. Davies, H. D. Ellis, and J. Shepherd. Face recognition accuracy as a function of mode of representation. Journal of Applied Psychology, 63(2):180, 1978.

[14] H. Dong and N. Gu. Asian face image database pf01. Technical report, Technical Report, Pohang University of Science and Technology, 2001.

[15] P. Ekman, T. S. Huang, T. J. Sejnowski, and J. C. Hager. Final report to nsf of the planning workshop on facial expression understanding. Human Interaction Laboratory, University of California, San Francisco, 378, 1993.

[16] A. Friedman et al. Framing pictures: The role of knowledge in automatized encoding and memory for gist. Journal of experimental psychology: General, 108(3):316–355, 1979.

[17] B. A. Golomb, D. Lawrence, and T. Sejnowski. Sexnet: A neural network identifies sex from human faces. Advances in neural information processing systems, 3:572–577, 1991.

[18] R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In Audio-and Video-Based Biometric Person Authentication, pages 10–18. Springer, 2003.

[19] G. Guo and G. Mu. A study of large-scale ethnicity estimation with gender and age variations. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, pages 79 –86, june 2010.

[20] M. A. Hearst, S. Dumais, E. Osman, J. Platt, and B. Scholkopf. Support vector machines. Intelligent Systems and their Applications, IEEE, 13(4):18–28, 1998.

[21] S. Hosoi, E. Takikawa, and M. Kawade. Ethnicity estimation with facial images. In Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, pages 195–200, 2004.

[22] Y. Hu, Z. Zhang, X. Xu, Y. Fu, and T. Huang. Building large scale 3d face database for face analysis. Multimedia Content Analysis and Mining, pages 343–350, 2007.

[23] G.-B. Huang and L. Chen. Enhanced random search based incremental extreme learning machine. Neurocomputing, 71(16):3460–3468, 2008.

[24] G.-B. Huang, L. Chen, and C.-K. Siew. Universal approximation using incremental constructive feedforward networks with random hidden nodes. Neural Networks, IEEE Transactions on, 17(4):879–892, 2006.

[25] G.-B. Huang, D. H. Wang, and Y. Lan. Extreme learning machines: a survey. International Journal of Machine Learning and Cybernetics, 2(2):107–122, 2011.

[26] A. Jain and J. Huang. Integrating independent components and linear discriminant analysis for gender classification. In Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, pages 159–163, 2004.

[27] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. Circuits and Systems for Video Technology, IEEE Transactions on, 14(1):4–20, 2004.

[28] M. Jones and P. Viola. Fast multi-view face detection. Mitsubishi Electric Research Lab TR-20003-96, 3, 2003.

[29] S. Khan, M. Nazir, S. Akram, and N. Riaz. Gender classification using image processing techniques: A survey. In Multitopic Conference (INMIC), 2011 IEEE 14th International, pages 25 –30, dec. 2011.

[30] P. Komarinski. Automated fingerprint identification systems (AFIS). Academic Press, 2005.

[31] X. Lu, A. K. Jain, et al. Ethnicity identification from face images. In Proceedings of SPIE, volume 5404, pages 114–123, 2004.

[32] J. Lyle, P. Miller, S. Pundlik, and D. Woodard. Soft biometric classification using periocular region features. In Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on, pages 1–7, 2010.

[33] A. M. Martinez. The ar face database. CVC Technical Report, 24, 1998.

[34] Y. Mehmood, M. Ishtiaq, M. Tariq, and M. Arfan Jaffar. Classifier ensemble optimization for gender classification using genetic algorithm. In Information and Emerging Technologies (ICIET), 2010 International Conference on, pages 1 –5, june 2010.

[35] B. Moghaddam and M.-H. Yang. Learning gender with support faces. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(5):707 –711, may 2002.

[36] Y. Moses, S. Ullman, S. Edelman, et al. Generalization to novel images in upright and inverted faces. PERCEPTION-LONDON-, 25:443–462, 1996.

[37] S. Mozaffari, H. Behravan, and R. Akbari. Gender classification using single frontal image per person: Combination of appearance and geometric based features. In Pattern Recognition (ICPR), 2010 20th International Conference on, pages 1192 – 1195, aug. 2010.

[38] M. Nazir, M. Ishtiaq, A. Batool, M. A. Jaffar, and A. M. Mirza. Feature selection for efficient gender classification. In WSEAS International conference, 2010.

[39] A. Oliva. Gist of the scene. Neurobiology of attention, 696:64, 2005.

[40] O. Ozbudak, M. Kirc, Y. Cakir, and E. Gunes. Effects of the facial and racial features on gender classification. In MELECON 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference, pages 26 –29, april 2010.

[41] K. Patterson and A. Baddeley. When face recognition fails. Journal of Experimental Psychology: Human Learning and Memory, 3(4):406–417, 1977.

[42] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The feret database and evaluation procedure for face-recognition algorithms. Image and vision computing, 16(5):295–306, 1998.

[43] K. Ricanek and E. Boone. The effect of normal adult aging on standard pca face recognition accuracy rates. In Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on, volume 4, pages 2018–2023 vol. 4, 31 2005-Aug. 4.

[44] K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, pages 341–345, 2006.

[45] K. Ricanek, Y. Wang, C. Chen, and S. Simmons. Generalized multi-ethnic face age-estimation. In Biometrics: Theory, Applications, and Systems, 2009. BTAS '09. IEEE 3rd International Conference on, pages 1–6, Sept.

[46] S. Roomi, S. Virasundarii, S. Selvamegala, S. Jeevanandham, and D. Hariharasudhan. Race classification based on facial features. In Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on, pages 54 –57, dec. 2011.

[47] A. L. Samuel. Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, 3(3):210–229, 1959.

[48] H. Schneiderman, N. Nechyba, and M. Sipe. Pittpatt, 2010.

[49] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 25(12):1615–1618, 2003.

[50] N. Subcomittee. Biometrics history @ONLINE, aug 2006.

[51] U. Tariq, Y. Hu, and T. Huang. Gender and ethnicity identification from silhouetted face profiles. In Image Processing (ICIP), 2009 16th IEEE International Conference on, pages 2441 –2444, nov. 2009.

[52] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 1, pages I–511. IEEE, 2001.

[53] P. Viola and M. J. Jones. Robust real-time face detection. International journal of computer vision, 57(2):137–154, 2004.

[54] R. W. Williams and K. Herrup. The control of neuron number. Annual review of neuroscience, 11(1):423–453, 1988.

[55] Z. Xu, L. Lu, and P. Shi. A hybrid approach to gender classification from face images. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, pages 1 –4, dec. 2008.

[56] Z. Yang and H. Ai. Demographic classification with local binary patterns. In Advances in Biometrics, pages 464–473. Springer, 2007.

Appendices A

Database Statistics

FG-NET				
Gender	**Race**			
	Greek			
Male	571			
Female	431			
Gender	**Age Range**			
	0-20	21-40	41-60	60+
Male	419	116	30	6
Female	291	106	32	2

Table 8.9: FG-NET Statistics

MORPH Album Non-Commercial					
Gender	**Race**				
	African	European	Asian	Hispanic	Other
Male	36,832	7,961	141	1,667	44
Female	5,757	2,598	13	102	19
Gender	**Age Range**				
	<20	20-29	30-39	40-49	50+
Male	6,638	14,016	12,447	10,062	3,482
Female	831	2,309	2,910	1,988	451
Gender	**Images per Subject**				
	1	2	3	4	5+
Male	373	2,350	3,606	1,975	3,155
Female	85	478	712	352	532

Table 8.10: MORPH Non-Commercial Statistics

MORPH Commercial					
Gender	**Race**				
	African	European	Asian	Hispanic	Other
Male	97,805	24,755	654	6,678	250
Female	17,855	7,694	86	476	60
Gender	**Age Range**				
	<20	20-29	30-39	40-49	50+
Male	19,217	44,417	30,142	25,623	10,743
Female	3,458	8,696	7,334	5,365	1,318
Gender	**Images per Subject**				
	1	2	3	4	5+
Male	10,782	5,751	3,395	2,298	6,806
Female	3,159	1,384	752	416	1,128

Table 8.11: MORPH Commercial Statistics

Pinellas County Sheriffs Office						
Gender	**Race**					
	Hispanic	Asian	Caucasian	Other	Native Amer.	African
Male	54,493	5,564	661,413	1,594	673	278,106
Female	6,541	1,519	213,648	314	171	74,637
Gender	**Age Range**					
	Unknown	1-20	21-40	41-60	61-80	81+
Male	16	116,161	574,912	291,099	19,221	434
Female	4	30,123	183,295	80,126	3,210	72

Table 8.12: Pinellas County Sheriffs Office Statistics

Dataset	**Unique Subjects**	**Total Subjects**
FGNET	82	1,002
MORPH	35,872	156,265
Pinellas	254,812	1,298,672

Table 8.13: Unique Subjects per Dataset

Dataset	**Race**					
	White Male	White Female	Black Male	Black Female	Other Male	Other Female
FGNET	0	0	0	0	48	34
MORPH	6,506	2,047	19,843	4,580	2,685	211
Pinellas	129,455	45,903	45,953	15,150	15,964	2,387

Table 8.14: Datasets Demographics Per Unique Subject

Appendices B

Training/Validation/Testing Set Statistics

MORPH 1,266	Gender		
	Male	Female	Total
Training	379	380	759
Validation	127	126	253
Testing	127	127	254
Total			**1,266**

Table 8.15: MORPH 1,266 Dataset Gender Distribution

MORPH 1,266	**Race**			
	White	Black	Other	**Total**
Training	251	251	257	759
Validation	82	89	82	253
Testing	89	82	83	254
Total				**1,266**

Table 8.16: MORPH 1,266 Dataset Race Distribution

MORPH 1,266	**Age Range**										
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91+	Total
Training	0	151	264	194	115	34	1	0	0	0	759
Validation	0	58	88	60	37	9	1	0	0	0	253
Testing	0	45	87	68	37	15	2	0	0	0	254
Total											**1266**

Table 8.17: MORPH 1,266 Dataset Age Distribution

MORPH 10,446	Gender		
	Male	Female	Total
Training	3,684	2,583	6,267
Validation	1,228	861	2,089
Testing	1,229	861	2,090
Total			**10,446**

Table 8.18: MORPH 10,446 Dataset Distribution

MORPH 10,446	**Race**			
	White	Black	Other	**Total**
Training	2,451	2,450	1,366	6,267
Validation	846	810	433	2,089
Testing	797	834	459	2,090
Total				**10,446**

Table 8.19: MORPH 10,466 Dataset Race Distribution

MORPH 10,446	**Age Range**										
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91+	Total
Training	0	1,203	2,297	1,471	989	266	39	2	0	0	6,267
Validation	0	372	741	510	347	106	13	0	0	0	2,089
Testing	0	365	780	519	342	75	8	1	0	0	2,090
Total											**10,446**

Table 8.20: MORPH 10,446 Dataset Age Distribution

Dataset	Male	Female	Total
MORPH 1,226 - Eval.	129,468	25,532	155,000
MORPH 10,446 - Eval.	123,960	21,860	145,820

Table 8.21: MORPH Evaluation Dataset Gender Distribution

Dataset	White	Black	Other	Total
MORPH 1,266 - Eval.	31,945	115,121	7,934	155,000
MORPH 10,446 - Eval.	28,273	111,449	6,098	145,820

Table 8.22: MORPH Evaluation Dataset Race Distribution

Dataset	**Age Range**									
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81+	Total
MORPH 1,266 - Eval.	0	28,796	50,182	36,802	29,170	9,222	780	48	0	155,000
MORPH 10,446 - Eval.	0	27,110	46,803	34,624	27,681	8,833	724	45	0	145,820

Table 8.23: MORPH Evaluation Dataset Age Distribution

Pinellas 14,332	Gender		
	Male	Female	Total
Training	4,296	4,297	8,593
Validation	1,432	1,432	2,864
Testing	1,433	1,432	2,865
Total			**14,322**

Table 8.24: Pinellas 14,332 Dataset Distribution

Pinellas 14,332	**Race**			
	White	Black	Other	**Total**
Training	2,874	2,841	2,878	8,593
Validation	981	982	901	2,864
Testing	919	951	995	2,865
Total				**14,322**

Table 8.25: Pinellas 14,332 Dataset Race Distribution

Pinellas 14,332	**Age Range**										
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91+	Total
Training	0	1,117	3,174	2,178	1,546	466	96	13	2	1	8,593
Validation	0	369	1,038	753	494	173	28	9	0	0	2,864
Testing	0	381	1,069	715	493	167	33	7	0	0	2,865
Total											**14,322**

Table 8.26: Pinellas 14,332 Dataset Age Distribution

Dataset	Male	Female	Total
Pinellas Eval.	994,681	289,669	1,284,350

Table 8.27: Pinellas Evaluation Dataset Gender Distribution

Dataset	White	Black	Other	Total
Pinellas Eval.	870,286	347,969	66,095	1,284,350

Table 8.28: Pinellas Evaluation Dataset Race Distribution

Dataset	**Age Range**										
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91+	Total
Pinellas Eval.	46	144391	414875	334405	274283	93602	18672	3573	429	74	1,284,350

Table 8.29: Pinellas Evaluation Dataset Age Distribution

Appendices C

PCA Statistics

FGNET			
	Male	Female	Total
LBP	62	38	100
HOG	60	40	100
GIST	53	47	100

Table 8.30: FGNET PCA Gender Distribution

MORPH			
	Male	Female	Total
LBP	13,039	2,587	15,626
HOG	12,959	2,667	15,626
GIST	13,069	2,557	15,626

Table 8.31: MORPH PCA Gender Distribution

Pinellas			
	Male	Female	Total
LBP	100,314	29,553	129,867
HOG	100,239	29,628	129,867
GIST	100,082	29,785	129,867

Table 8.32: Pinellas PCA Gender Distribution

FGNET	**Race**						
	White Male	White Female	Black Male	Black Female	Other Male	Other Female	Total
LBP	0	0	0	0	62	38	100
HOG	0	0	0	0	60	40	100
GIST	0	0	0	0	53	47	100

Table 8.33: FGNET PCA Race Distribution

MORPH	**Race**						
	White Male	White Female	Black Male	Black Female	Other Male	Other Female	Total
LBP	2,463	759	9,757	1,762	819	66	15,626
HOG	2,348	761	9,860	1,837	751	69	15,626
GIST	2,548	722	9,710	1,774	811	61	15,626

Table 8.34: MORPH PCA Race Distribution

Pinellas	**Race**						
	White Male	White Female	Black Male	Black Female	Other Male	Other Female	Total
LBP	66,163	21,366	27,860	7,319	6,291	868	129,867
HOG	66,166	21,308	27,860	7,479	6,213	841	129,867
GIST	66,161	21,415	27,707	7,527	6,214	843	129,867

Table 8.35: Pinellas PCA Race Distribution

FGNET	**Age Range**										
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91+	Total
LBP	40	29	18	6	5	0	2	0	0	0	100
HOG	39	33	14	7	3	3	1	0	0	0	100
GIST	28	36	17	9	8	1	1	0	0	0	100

Table 8.36: FGNET PCA Age Distribution

MORPH	**Age Range**										
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91+	Total
LBP	0	2,925	5,093	3,705	2,927	907	64	5	0	0	15,626
HOG	0	2,912	5,102	3,620	2,937	963	85	7	0	0	15,626
GIST	0	2,888	5,184	3,608	2,957	903	78	8	0	0	15,626

Table 8.37: MORPH PCA Age Distribution

Pinellas	**Age Range**										
	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91+	Total
LBP	4	14,803	42,214	33,765	27,460	9,408	1,814	344	49	6	129,867
HOG	3	14,553	42,270	33,609	27,795	9,372	1,880	340	36	9	129,867
GIST	3	14,708	41,799	33,944	27,777	9,317	1,883	385	45	6	129,867

Table 8.38: Pinellas PCA Age Distribution

Appendices D

SVM Results

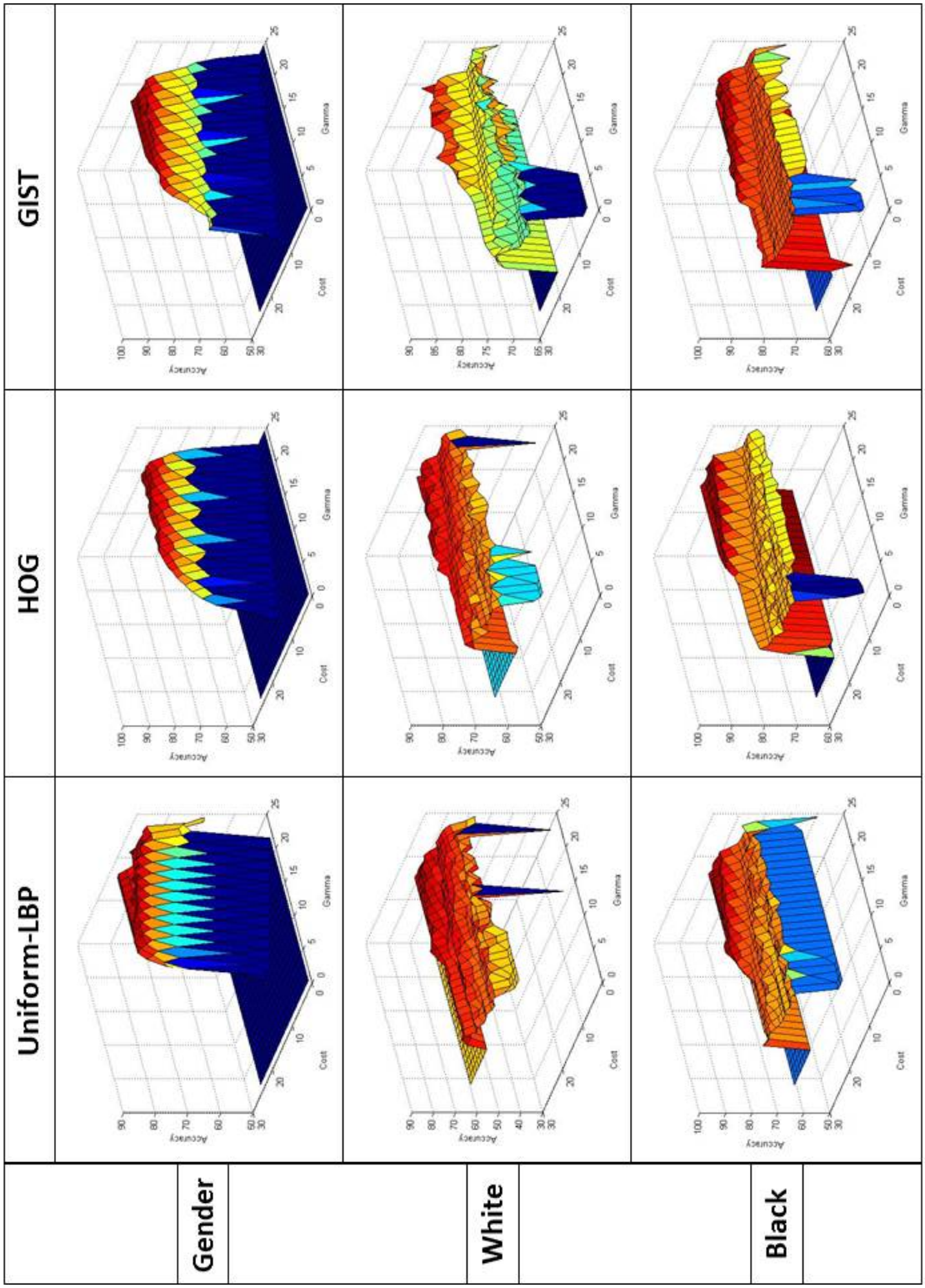

Figure 8.27: MORPH 1,266 Grid searches

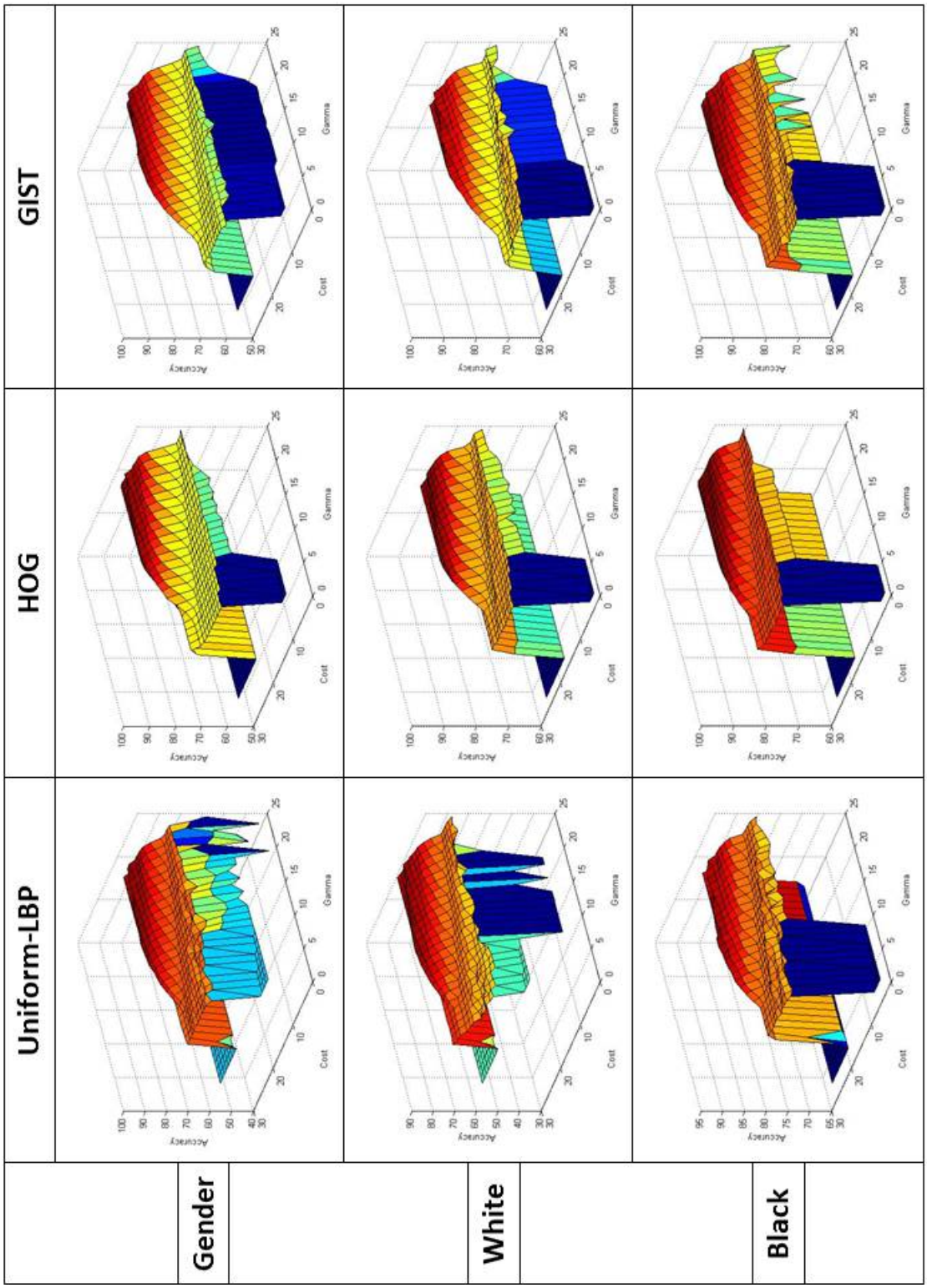

Figure 8.28: MORPH 10,446 Grid searches

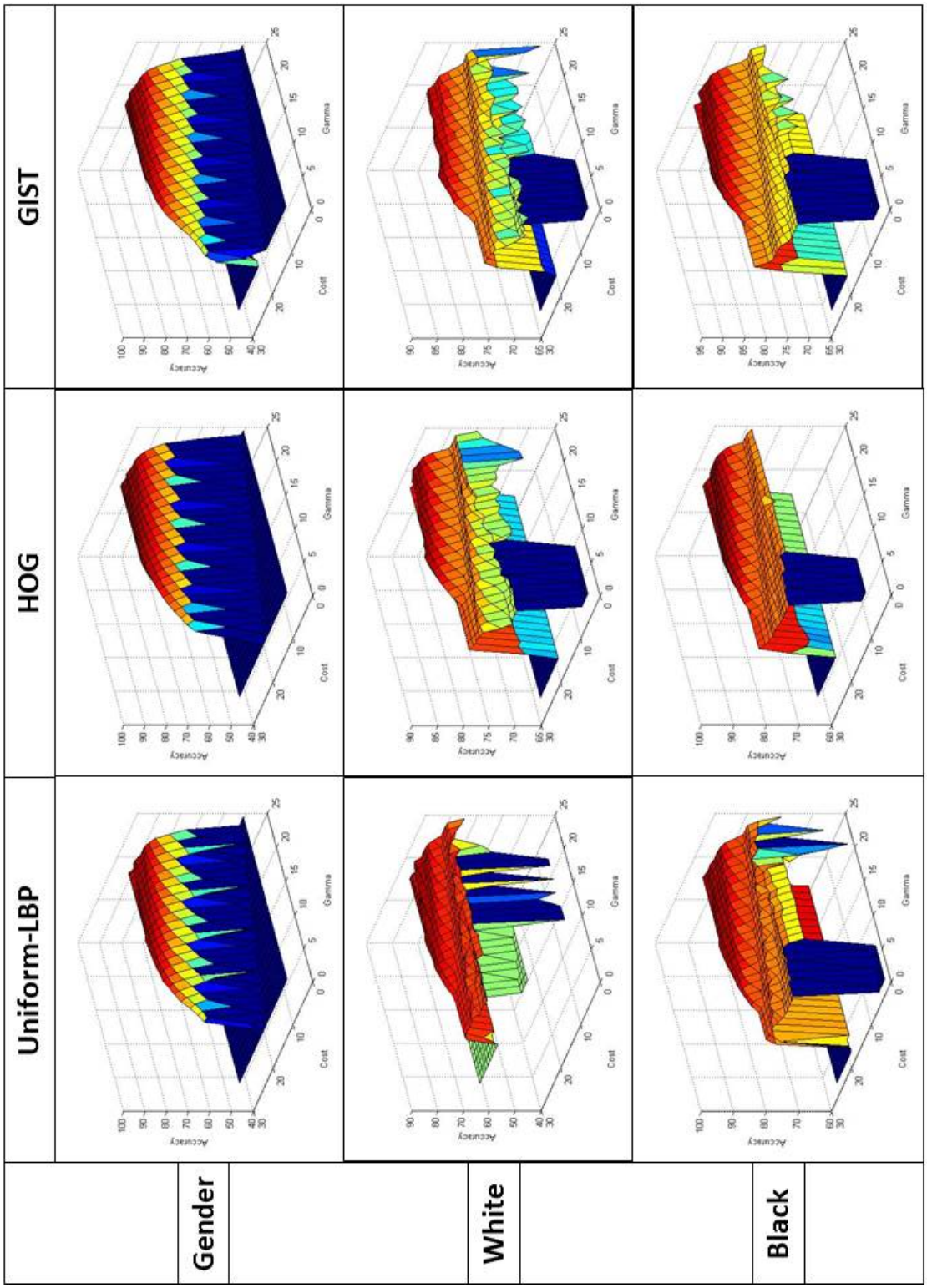

Figure 8.29: Pinellas 14,332 Grid searches

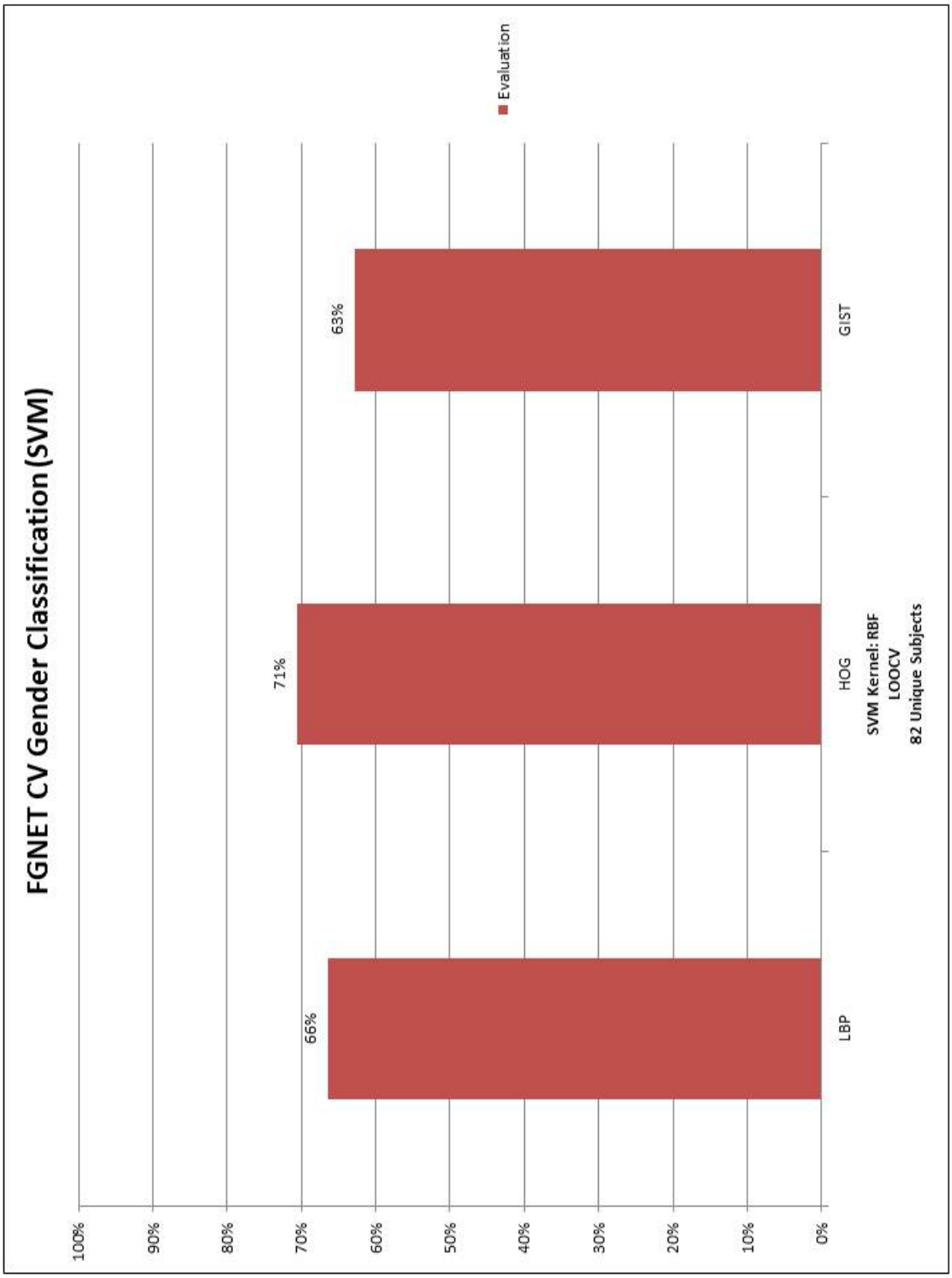

Figure 8.30: FGNET CV Gender Results

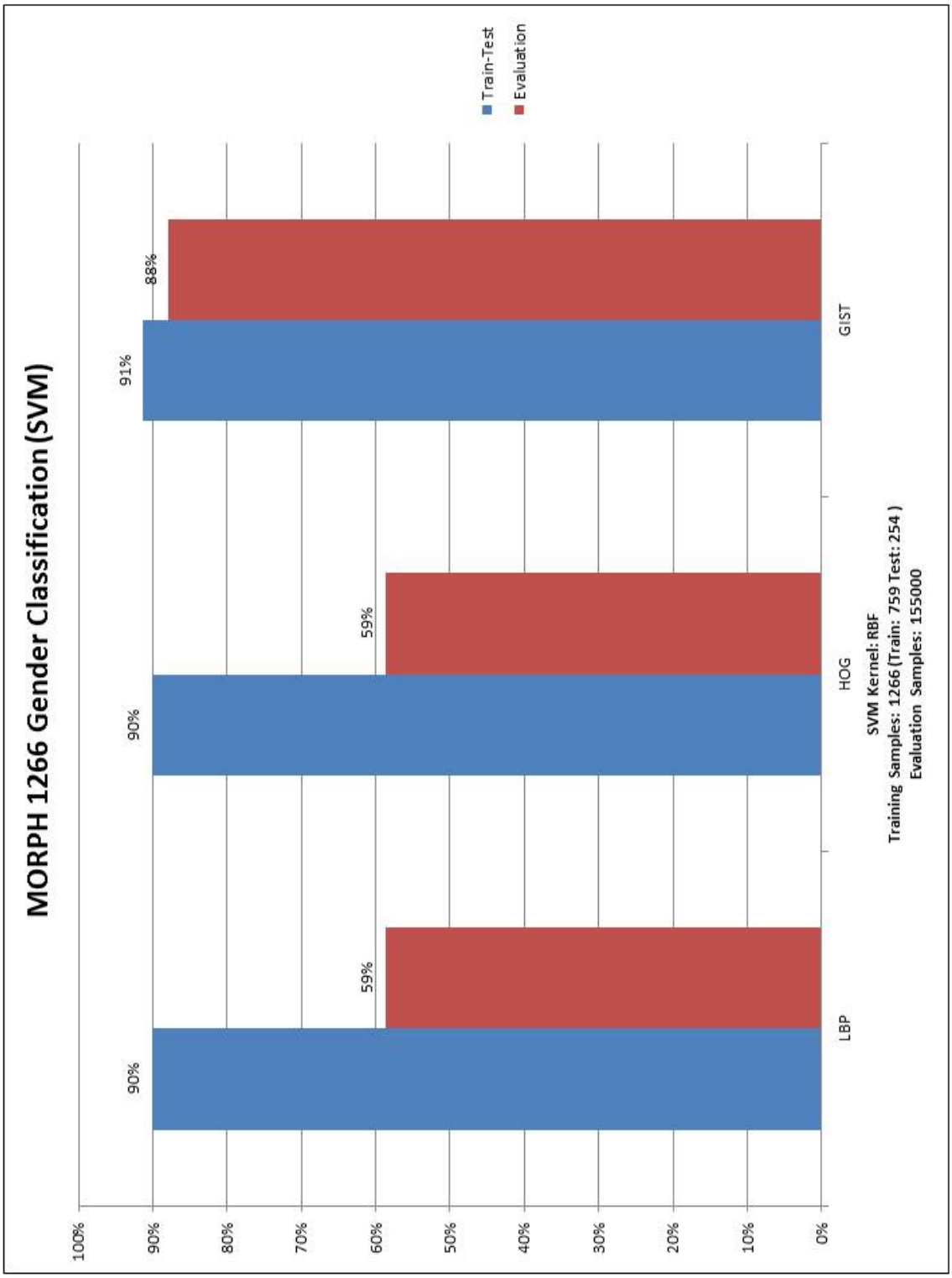

Figure 8.31: MORPH 1,266 Gender Results

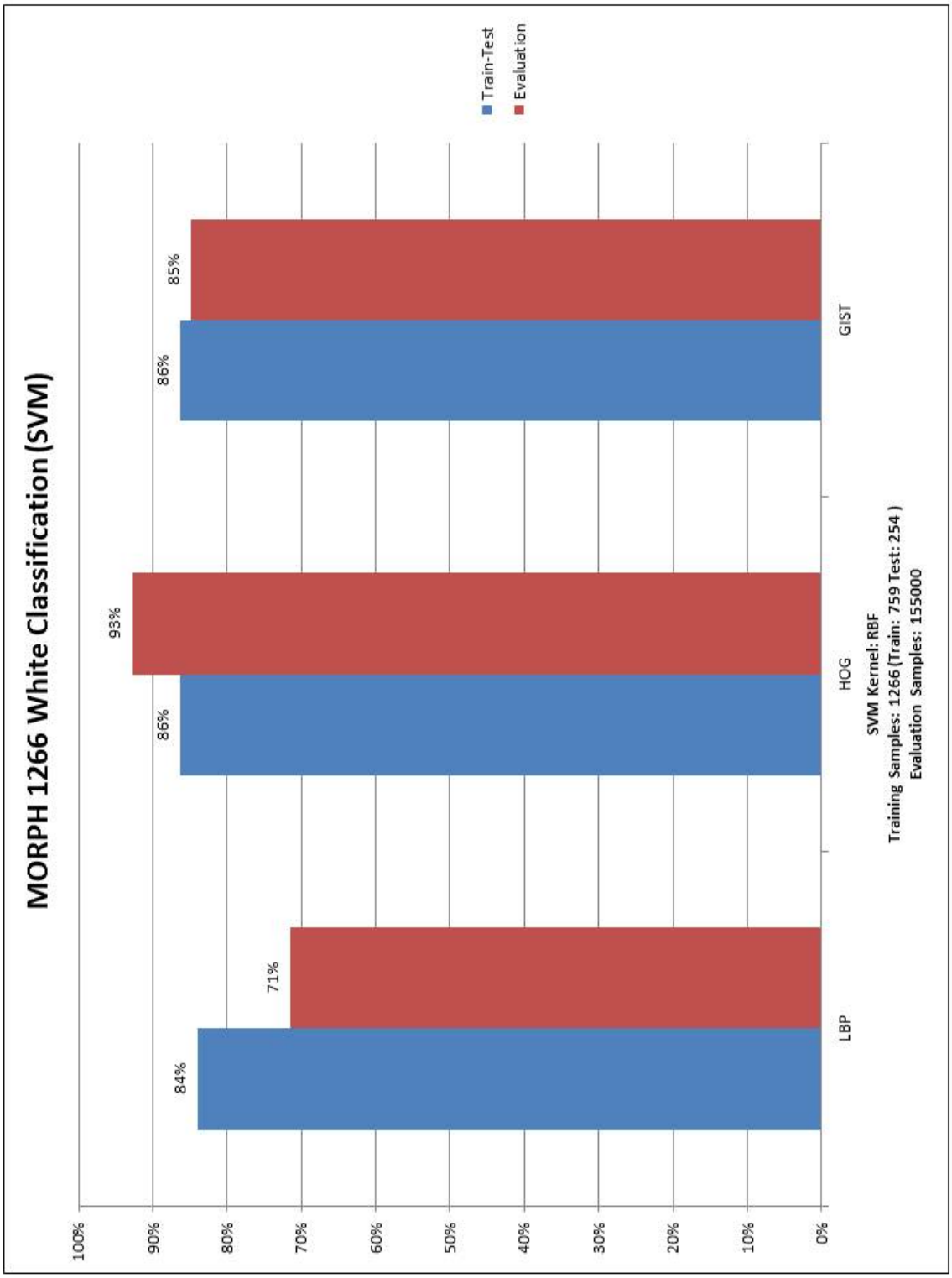

Figure 8.32: MORPH 1,266 White Results

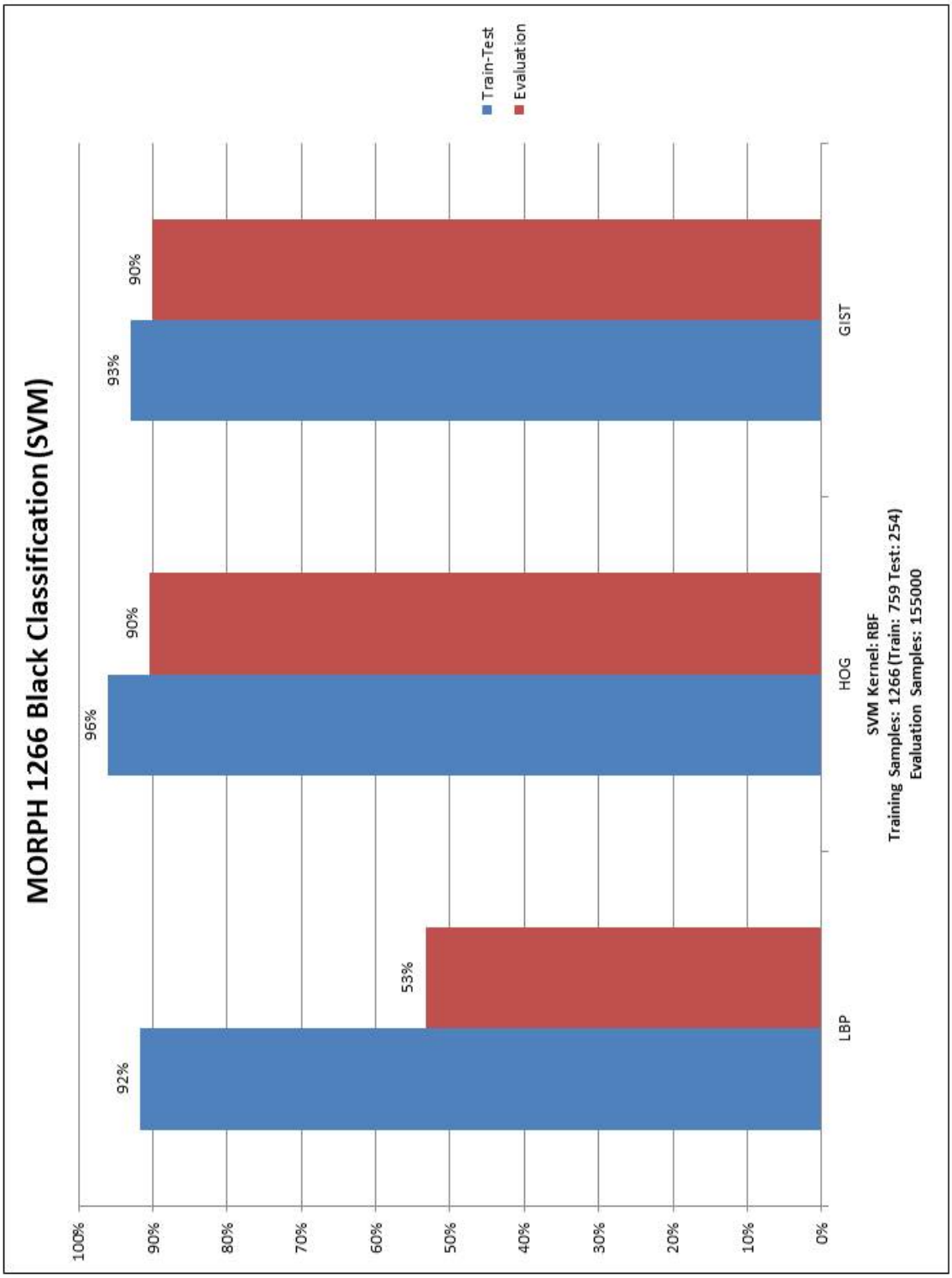

Figure 8.33: MORPH 1,266 Black Results

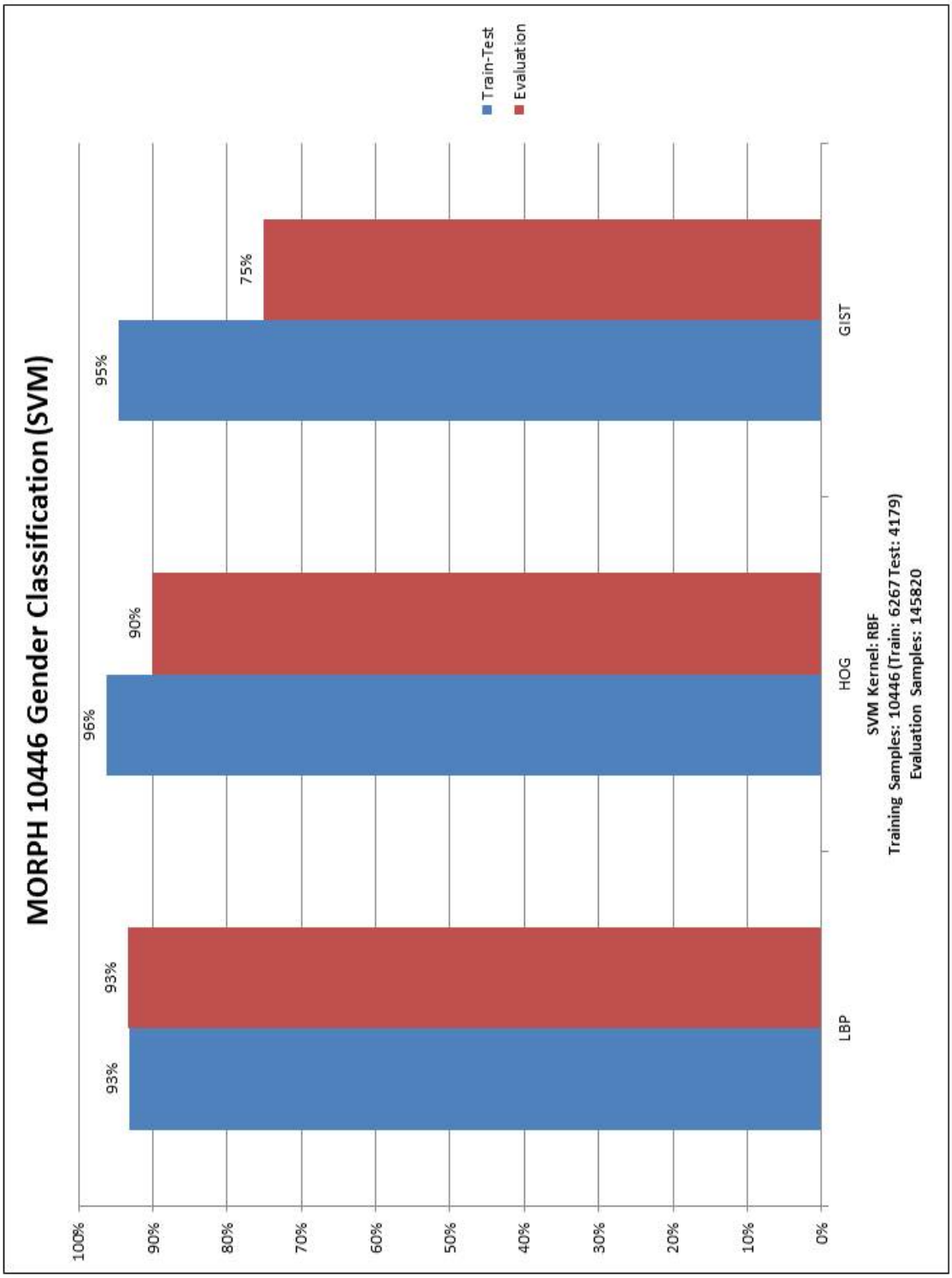

Figure 8.34: MORPH 10,446 Gender Results

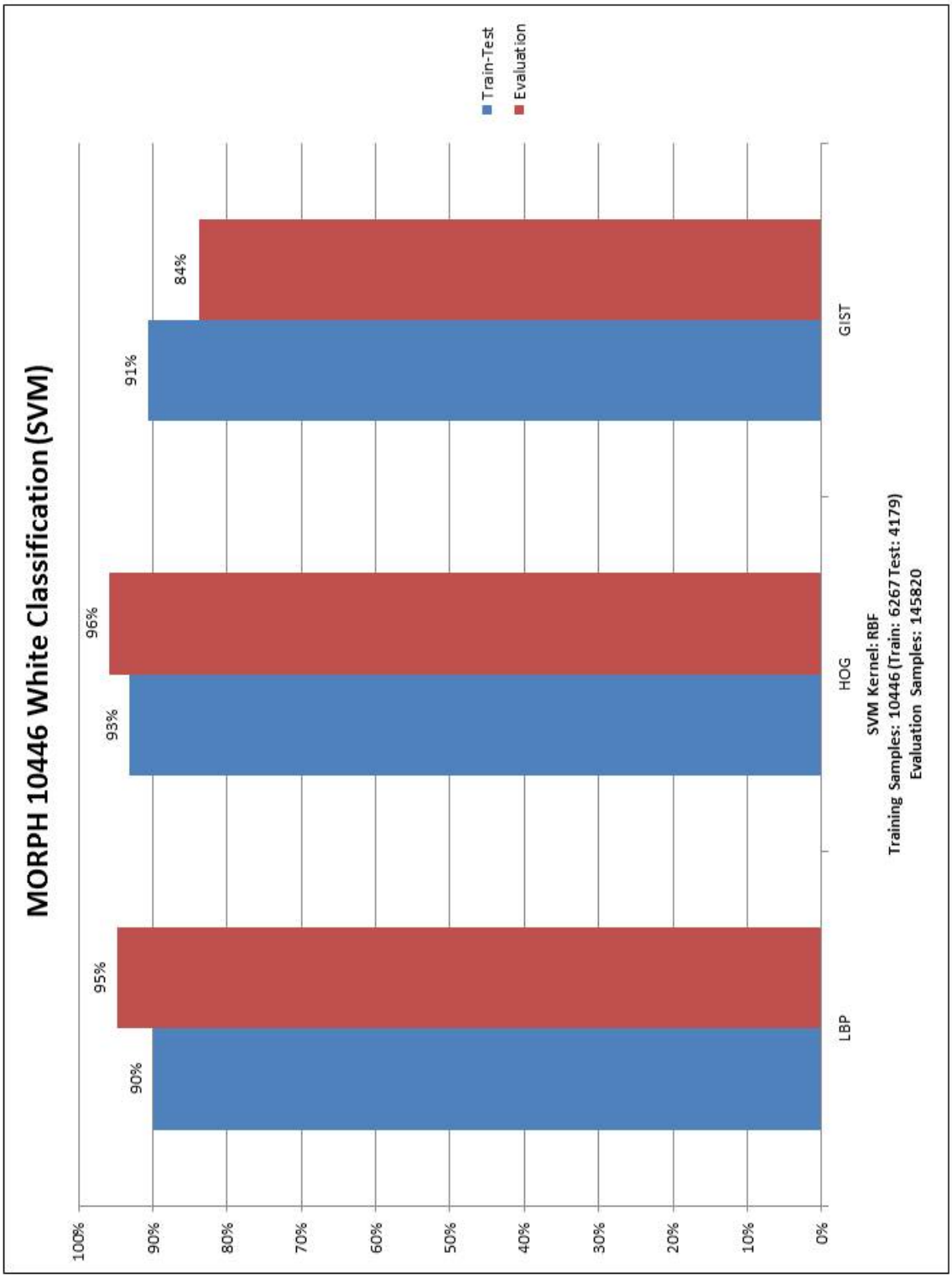

Figure 8.35: MORPH 10,446 White Results

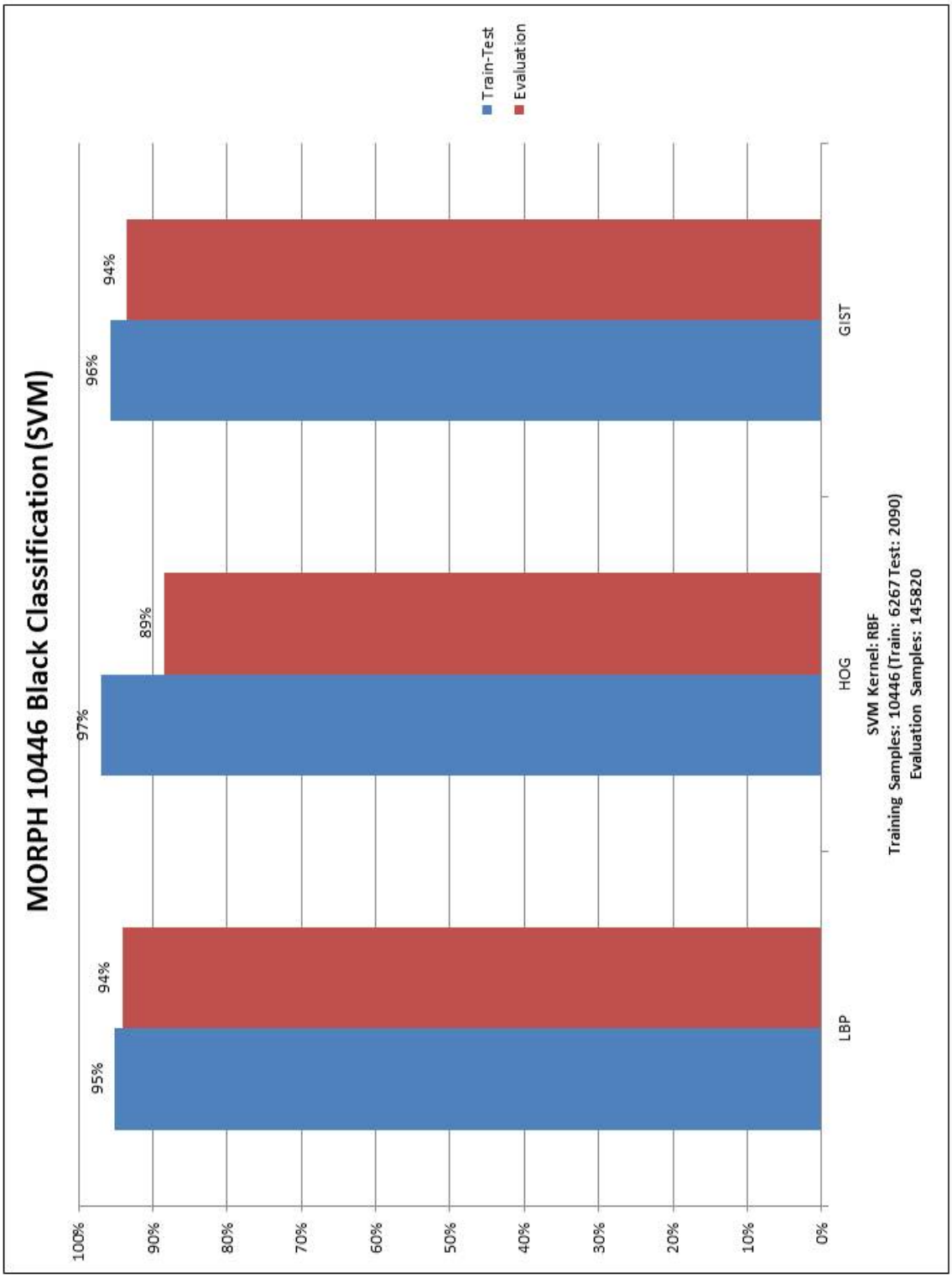

Figure 8.36: MORPH 10,446 Black Results

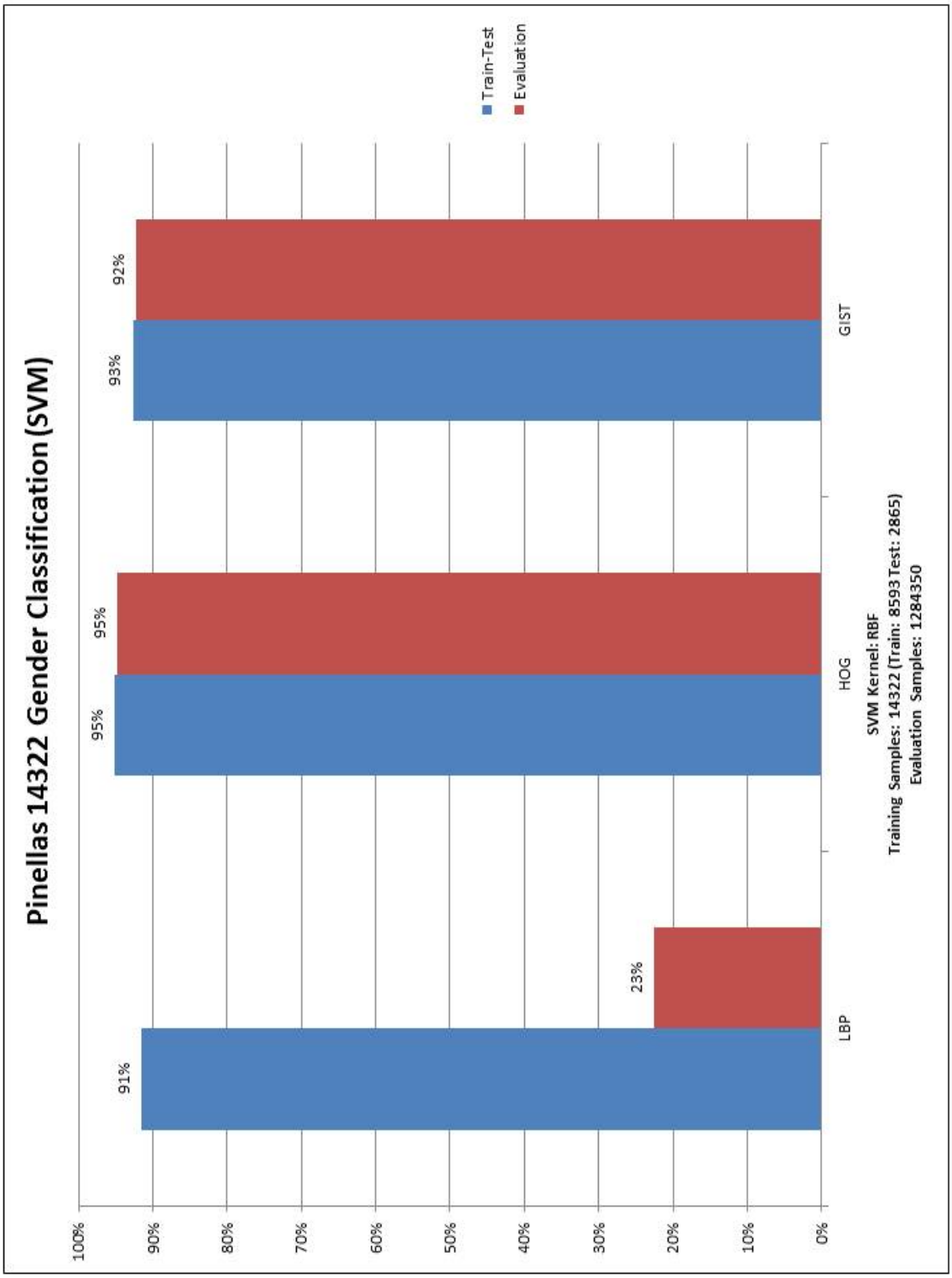

Figure 8.37: Pinellas 14,322 Gender Results

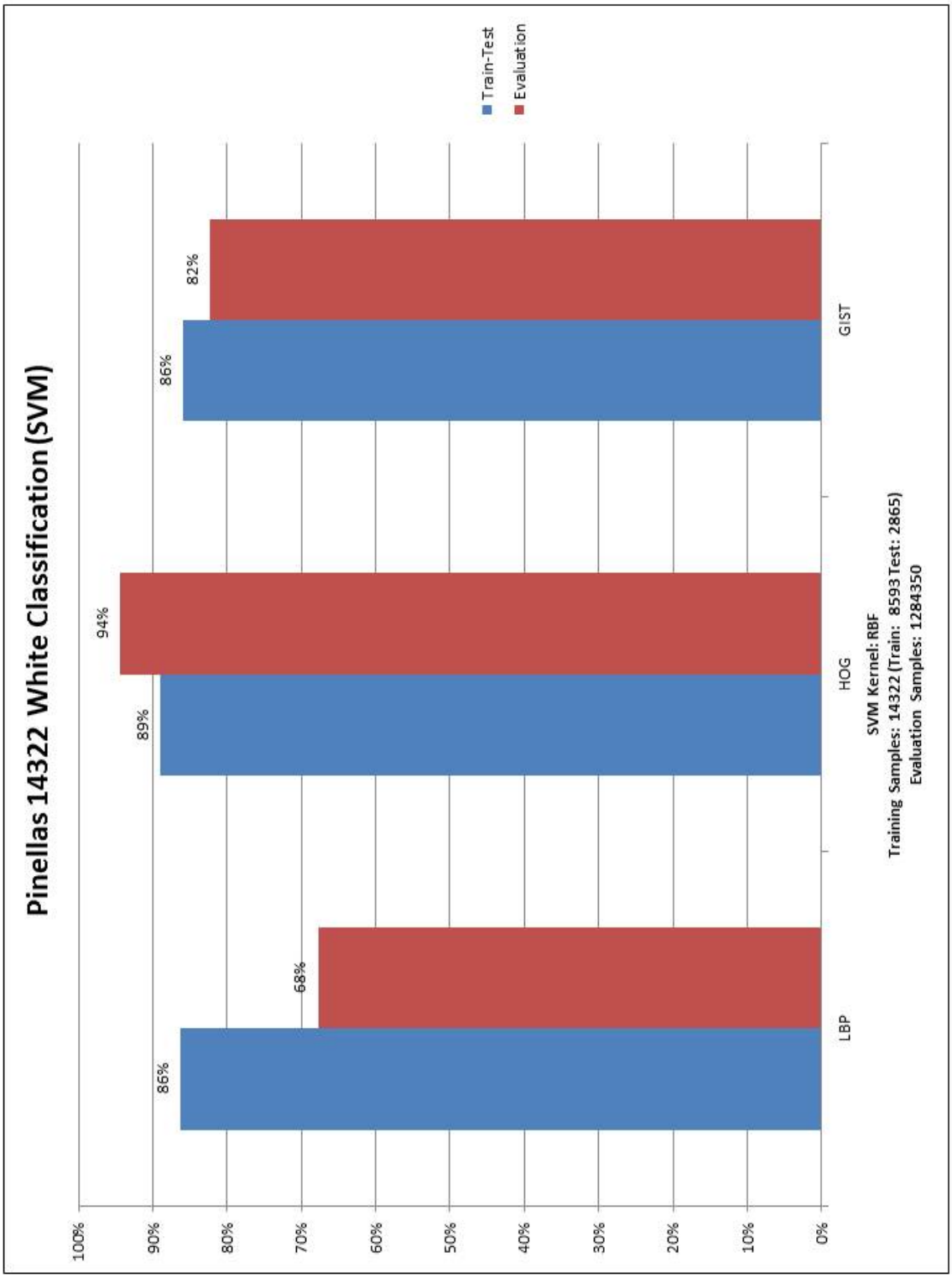

Figure 8.38: Pinellas 14,322 White Results

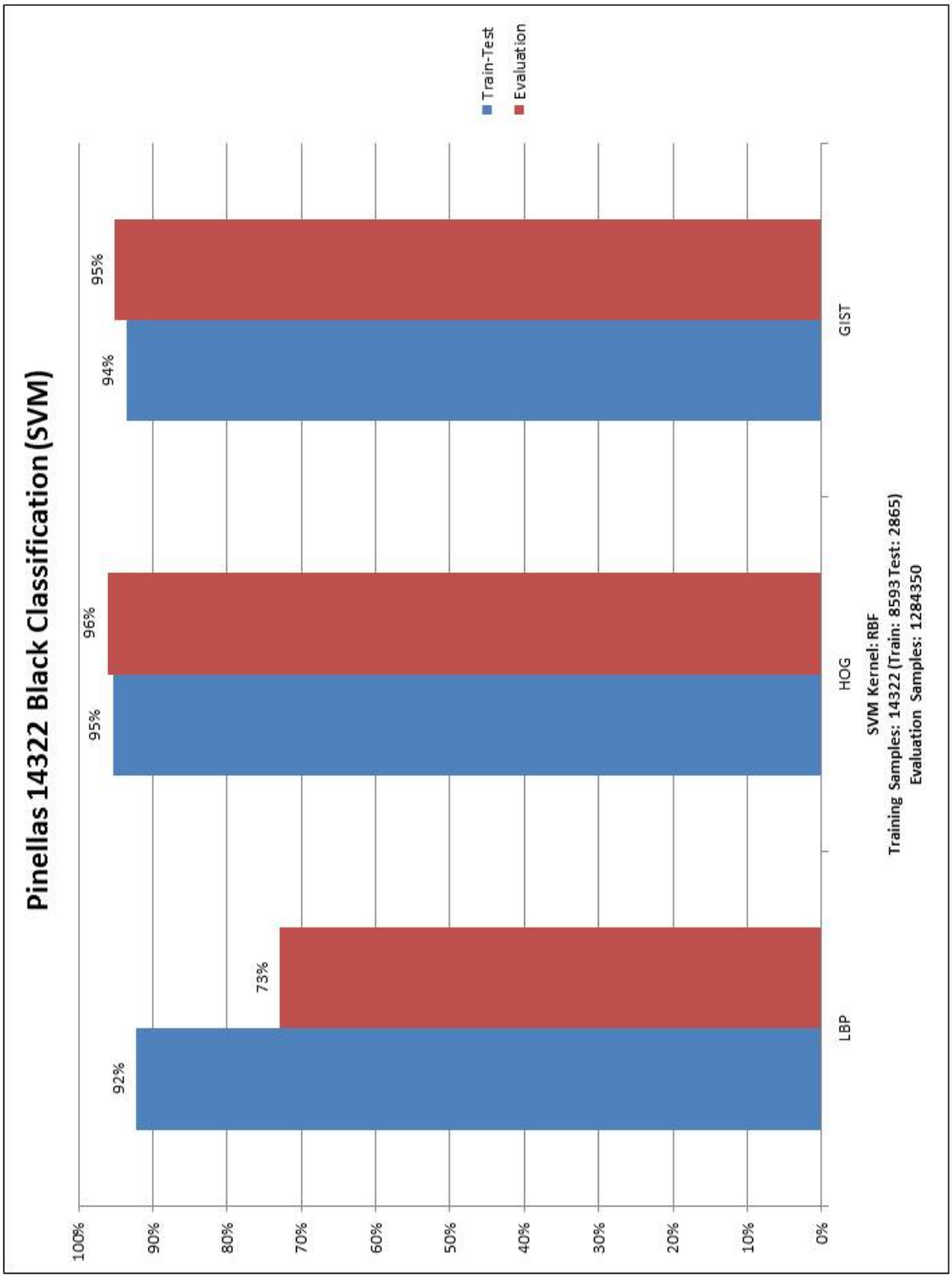

Figure 8.39: Pinellas 14,322 Black Results

Appendices E

ANN Results

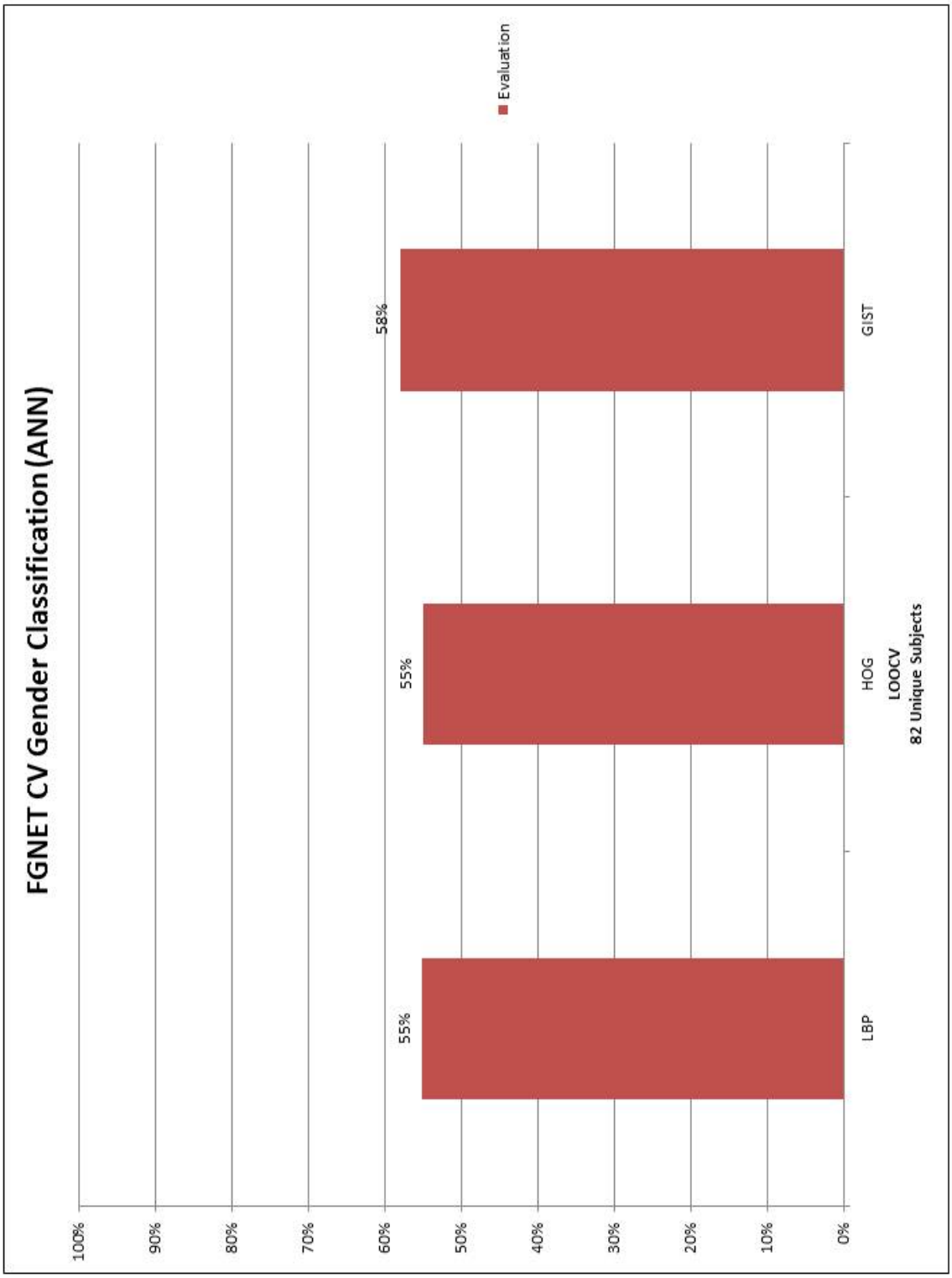

Figure 8.40: FGNET CV Gender Results

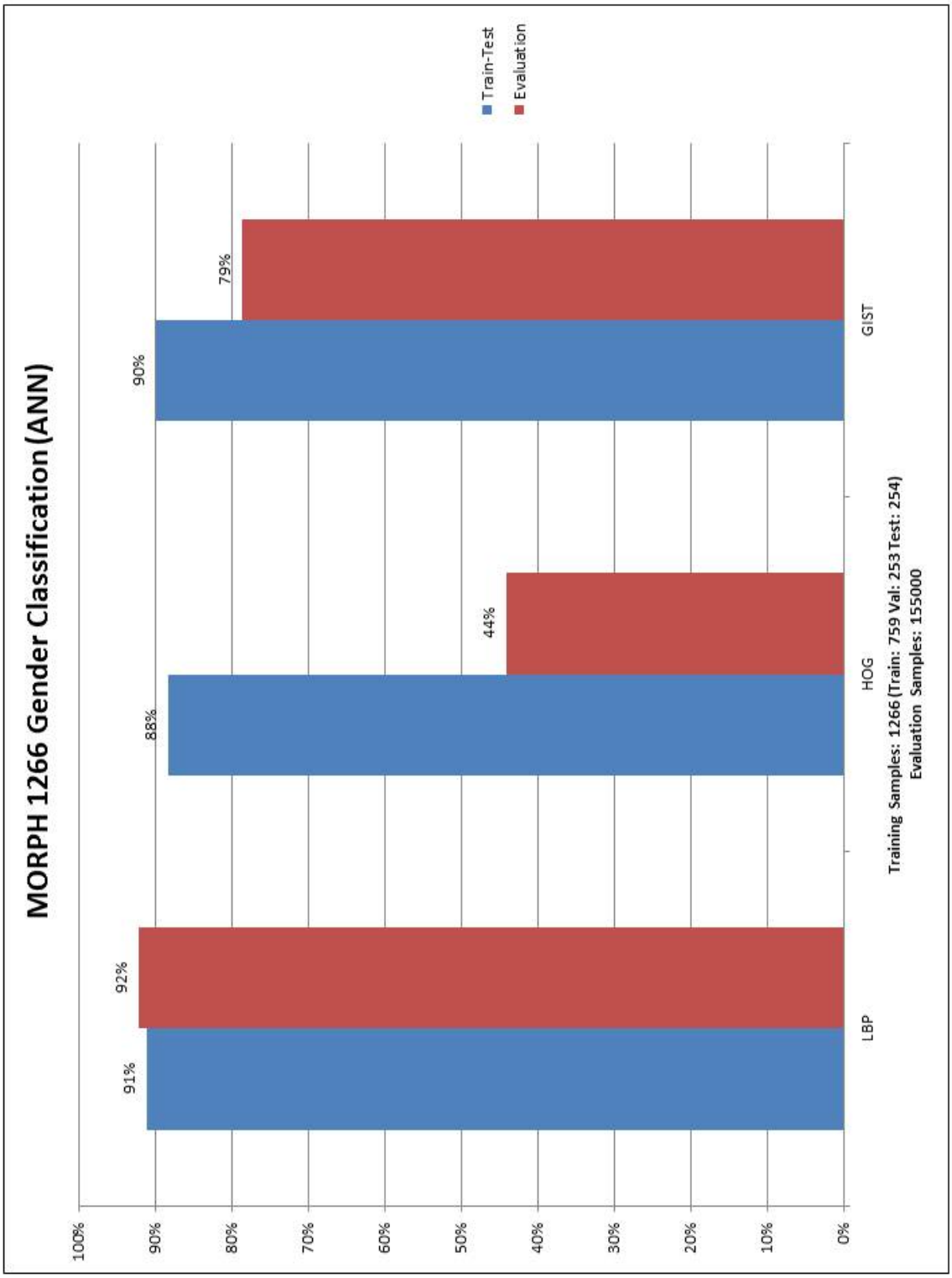

Figure 8.41: MORPH 1,266 Gender Results

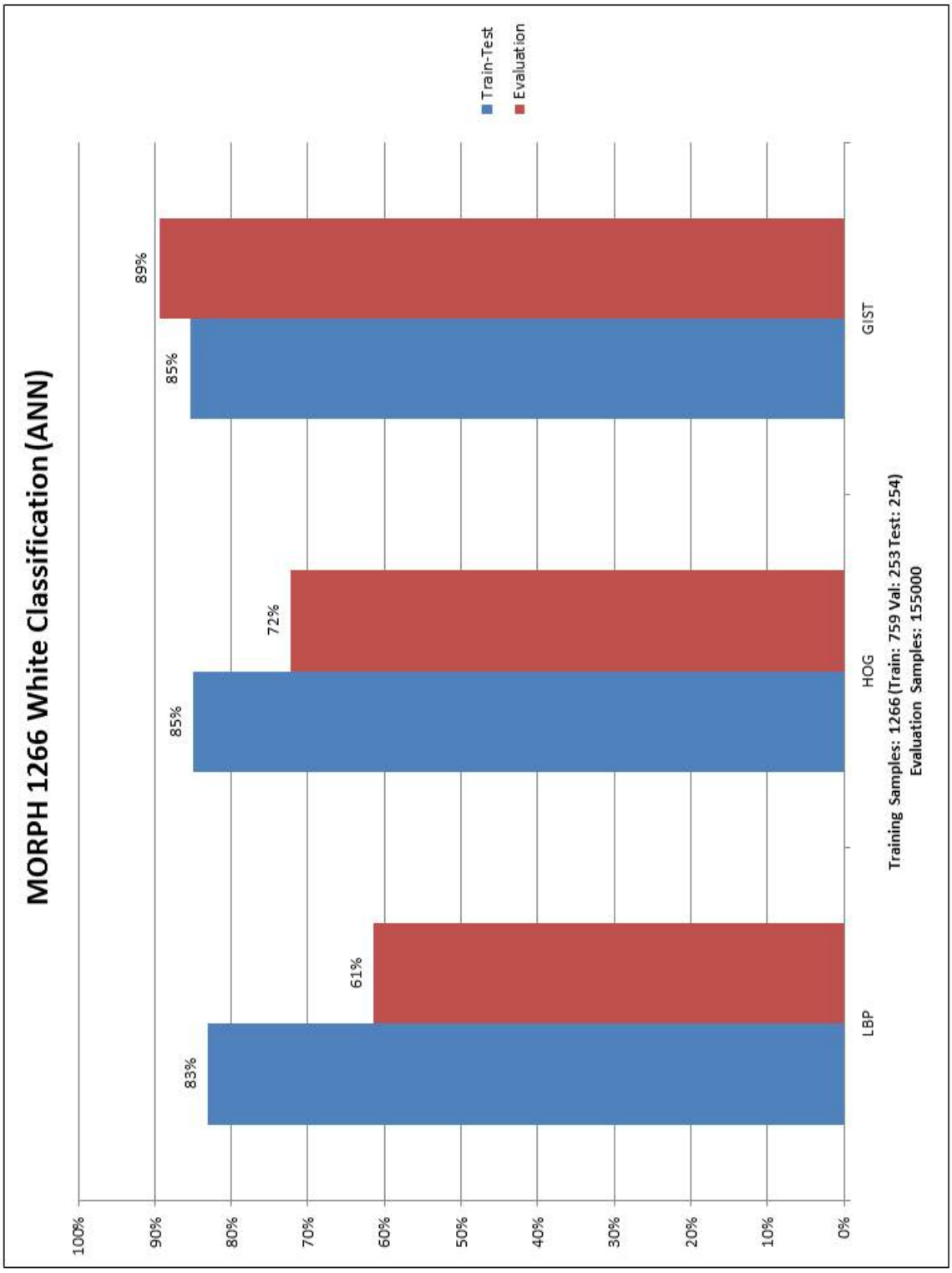

Figure 8.42: MORPH 1,266 White Results

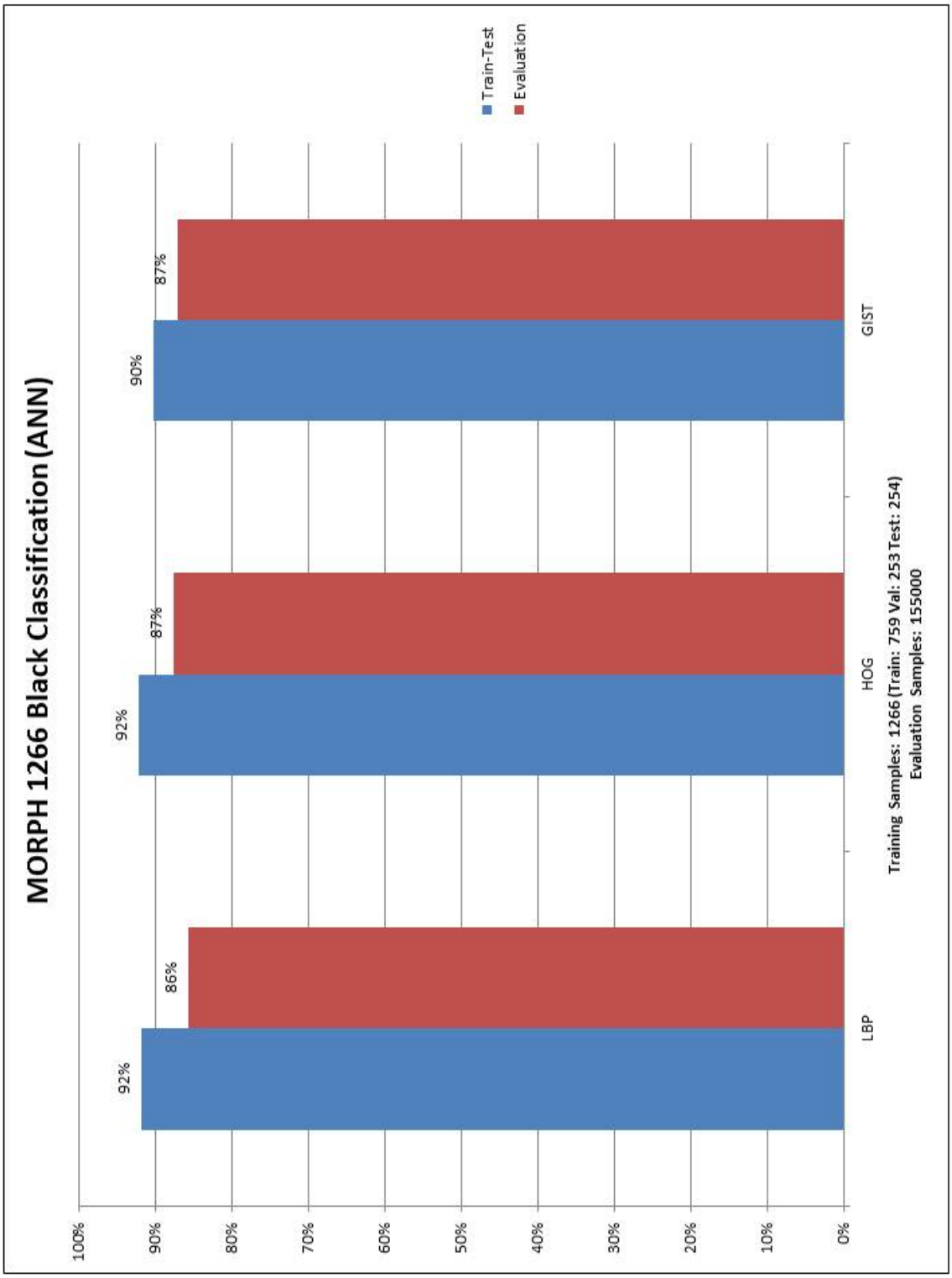

Figure 8.43: MORPH 1,266 Black Results

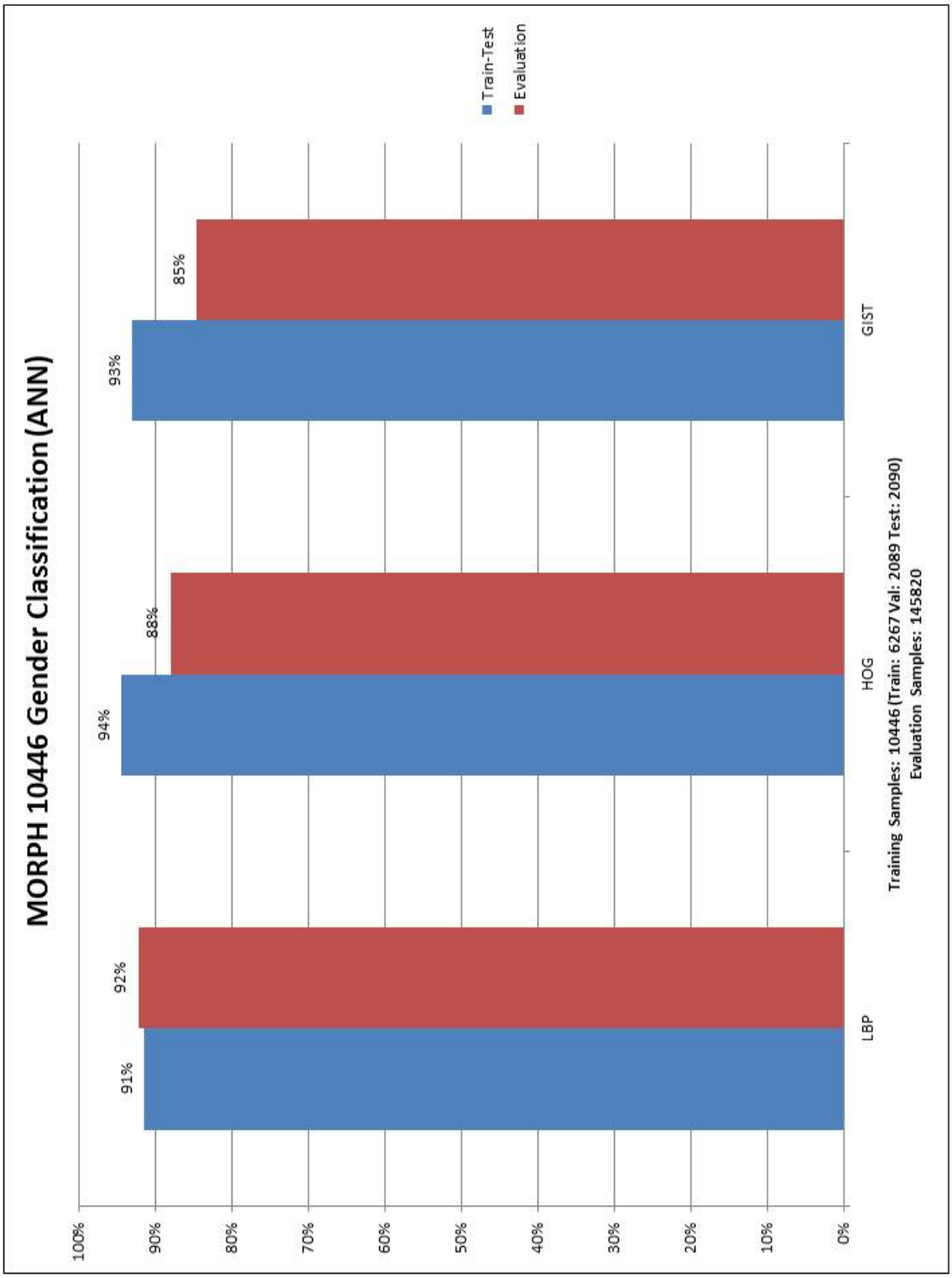

Figure 8.44: MORPH 10,446 Gender Results

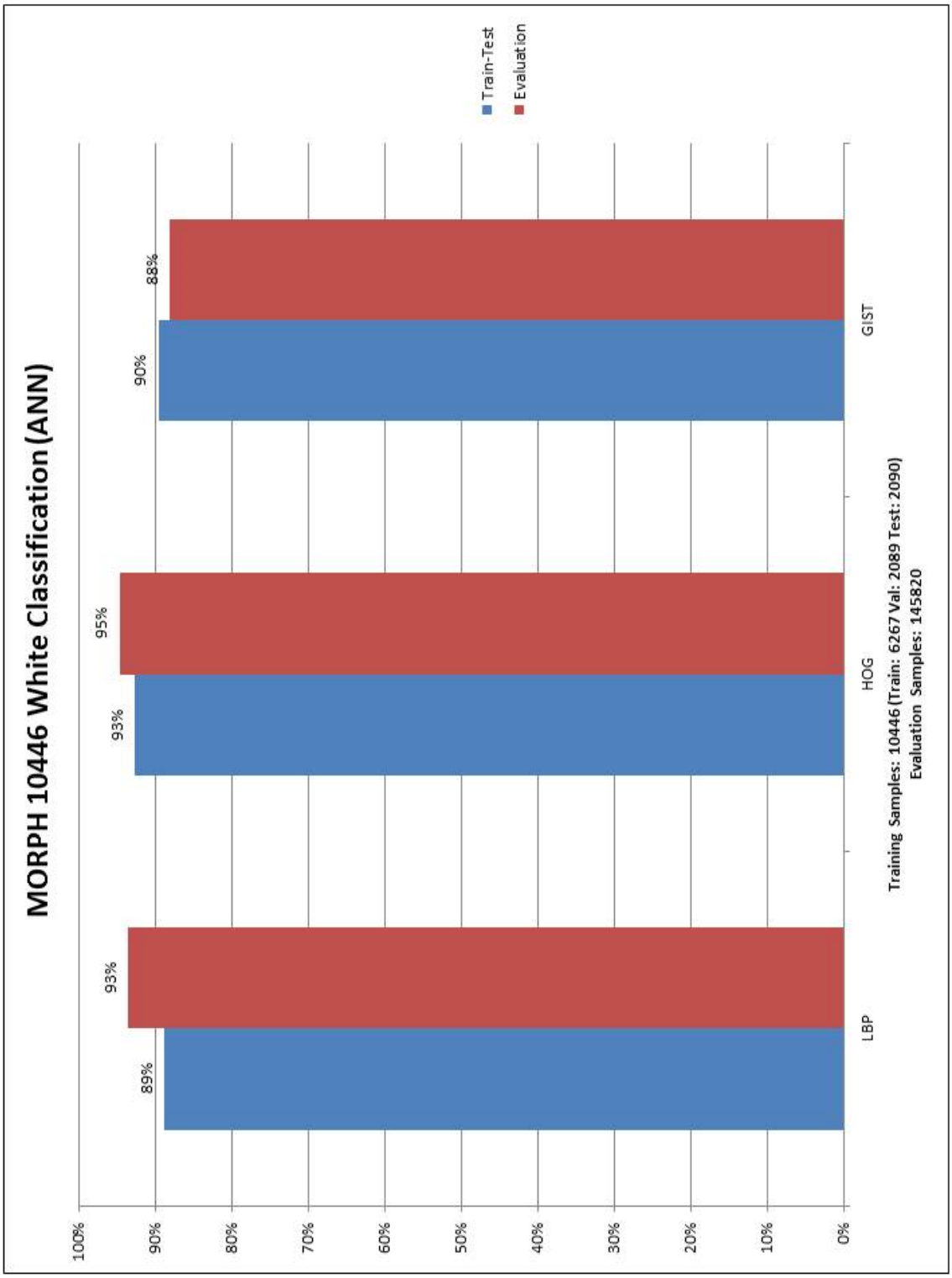

Figure 8.45: MORPH 10,446 White Results

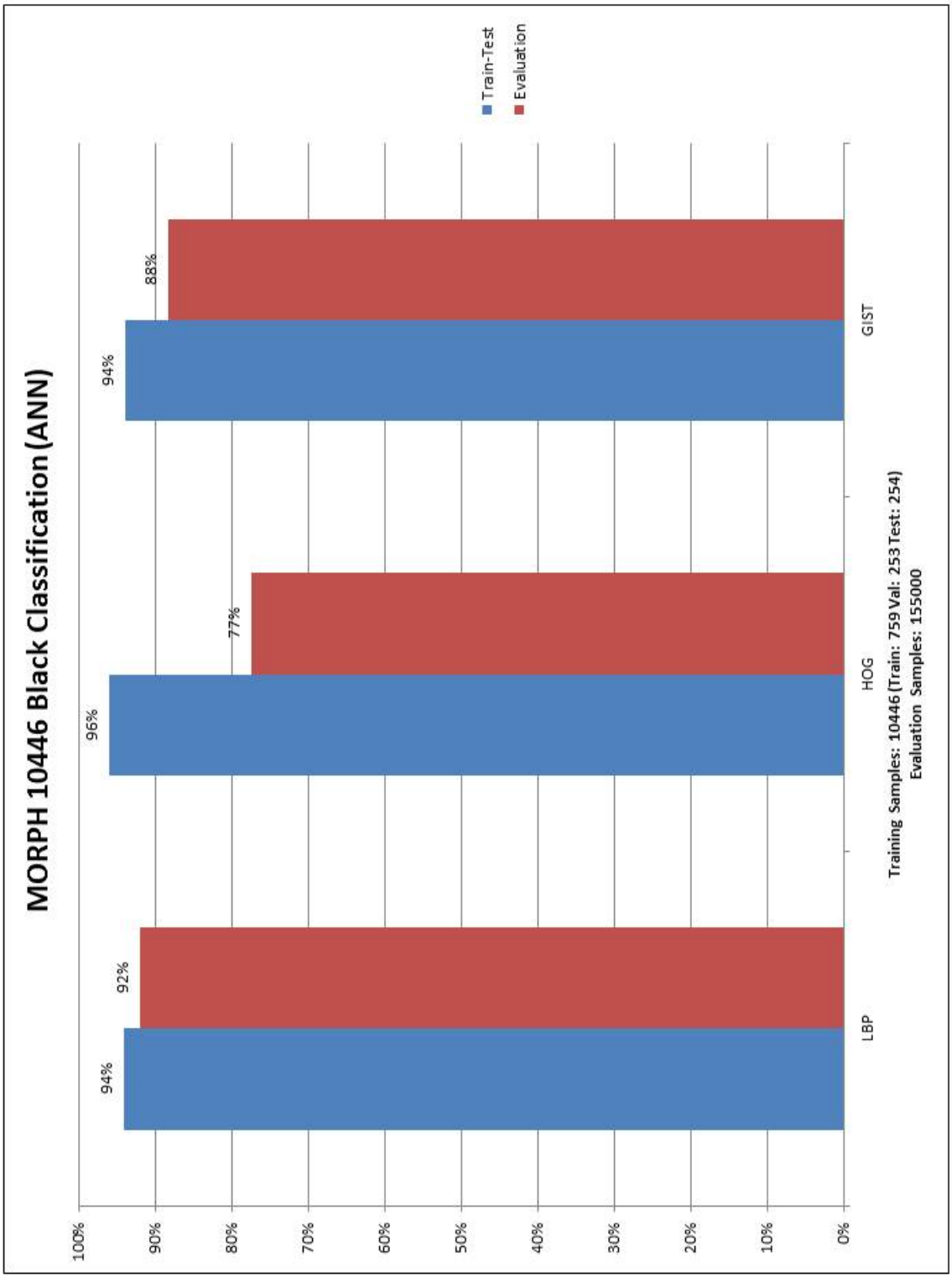

Figure 8.46: MORPH 10,446 Black Results

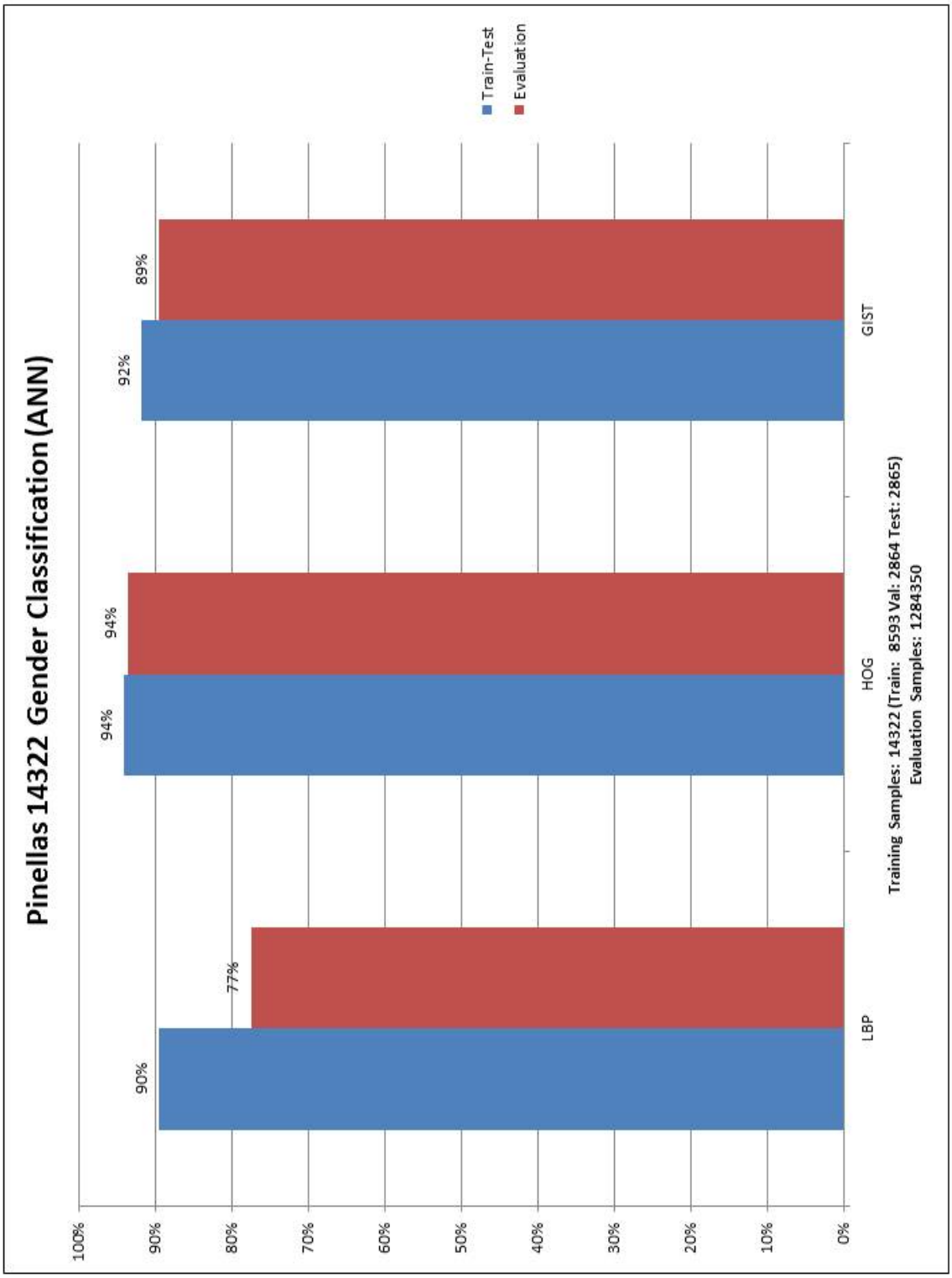

Figure 8.47: Pinellas 14,322 Gender Results

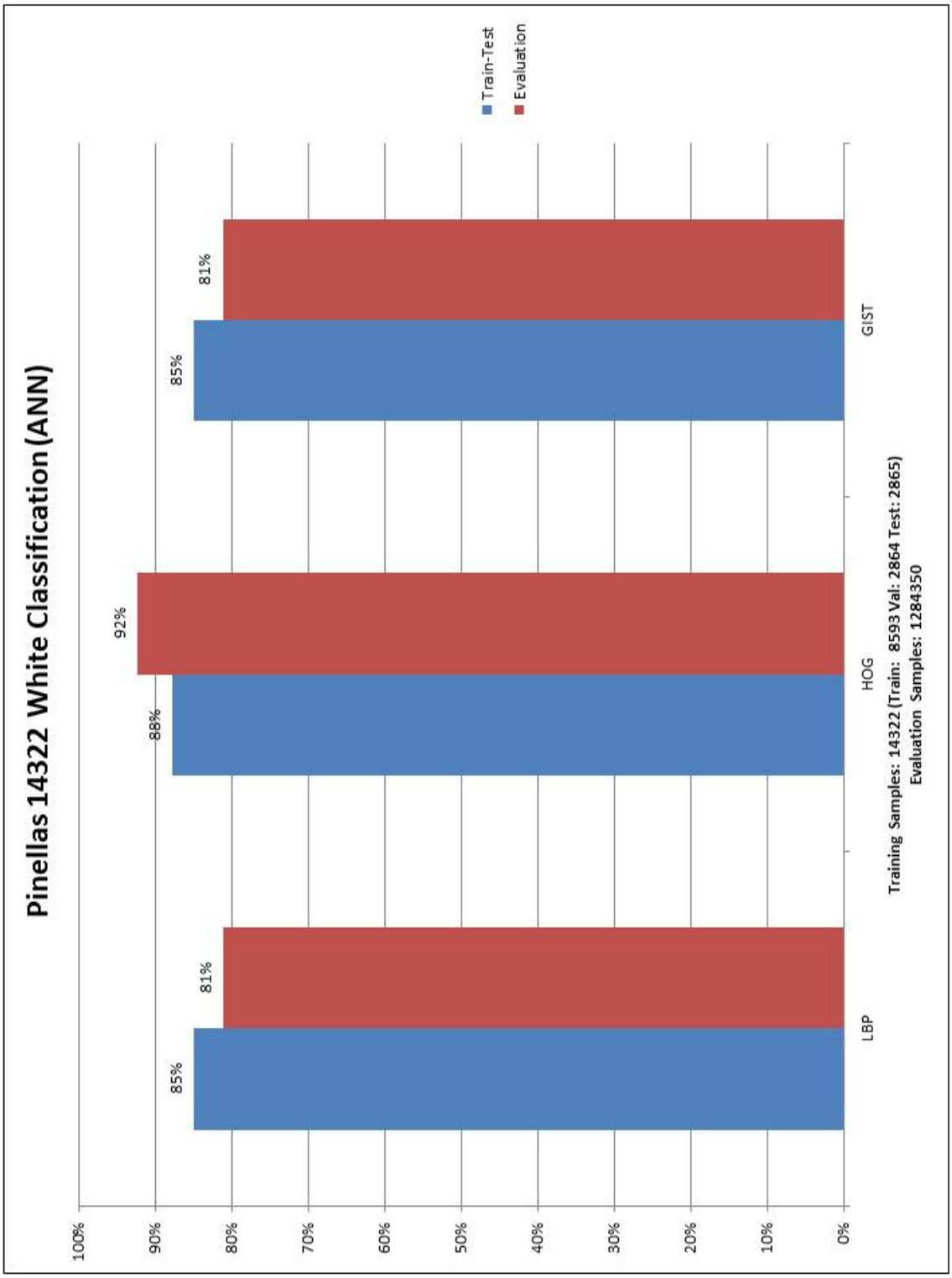

Figure 8.48: Pinellas 14,322 White Results

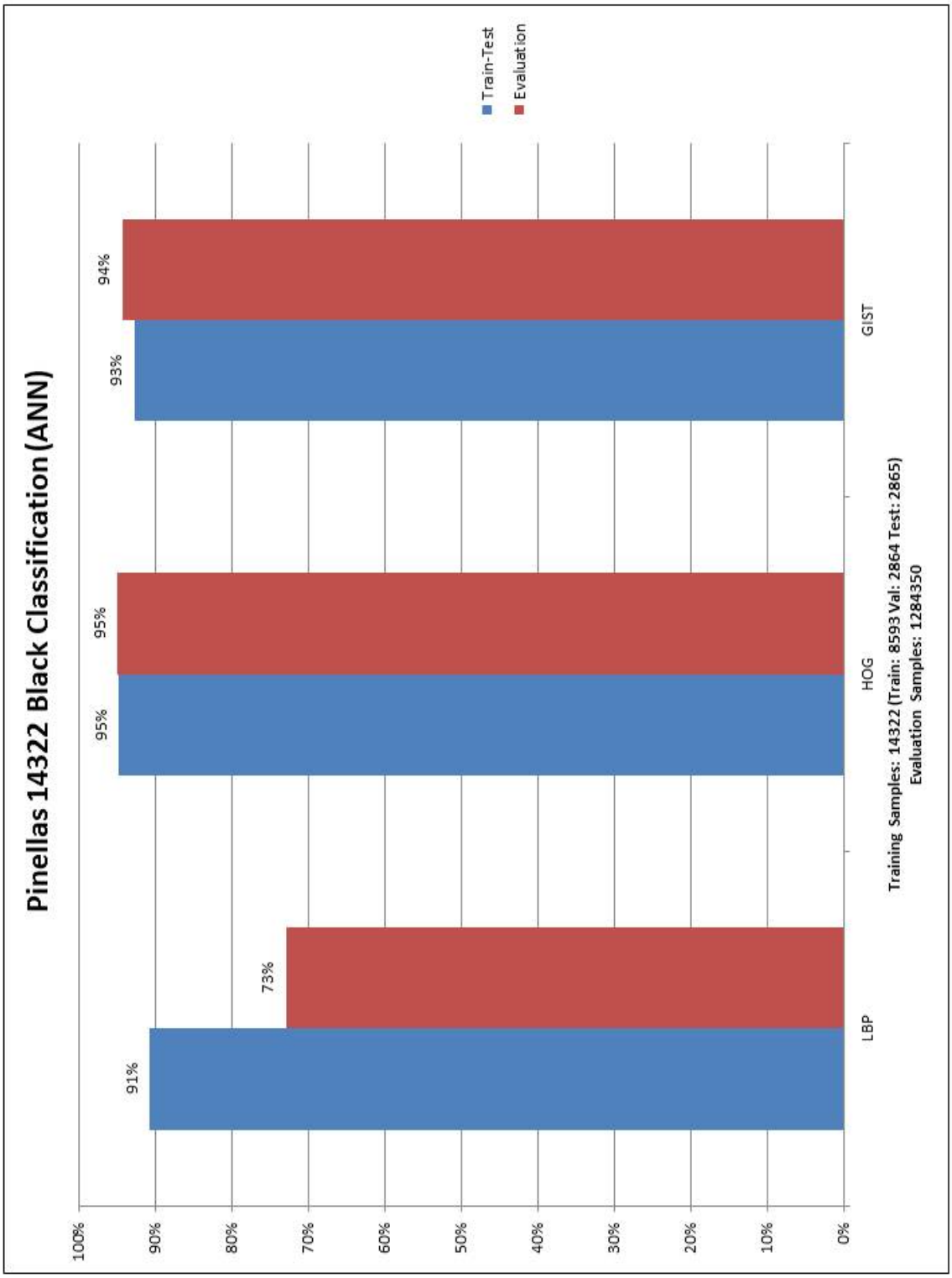

Figure 8.49: Pinellas 14,322 Black Results

Appendices F

ELM Results

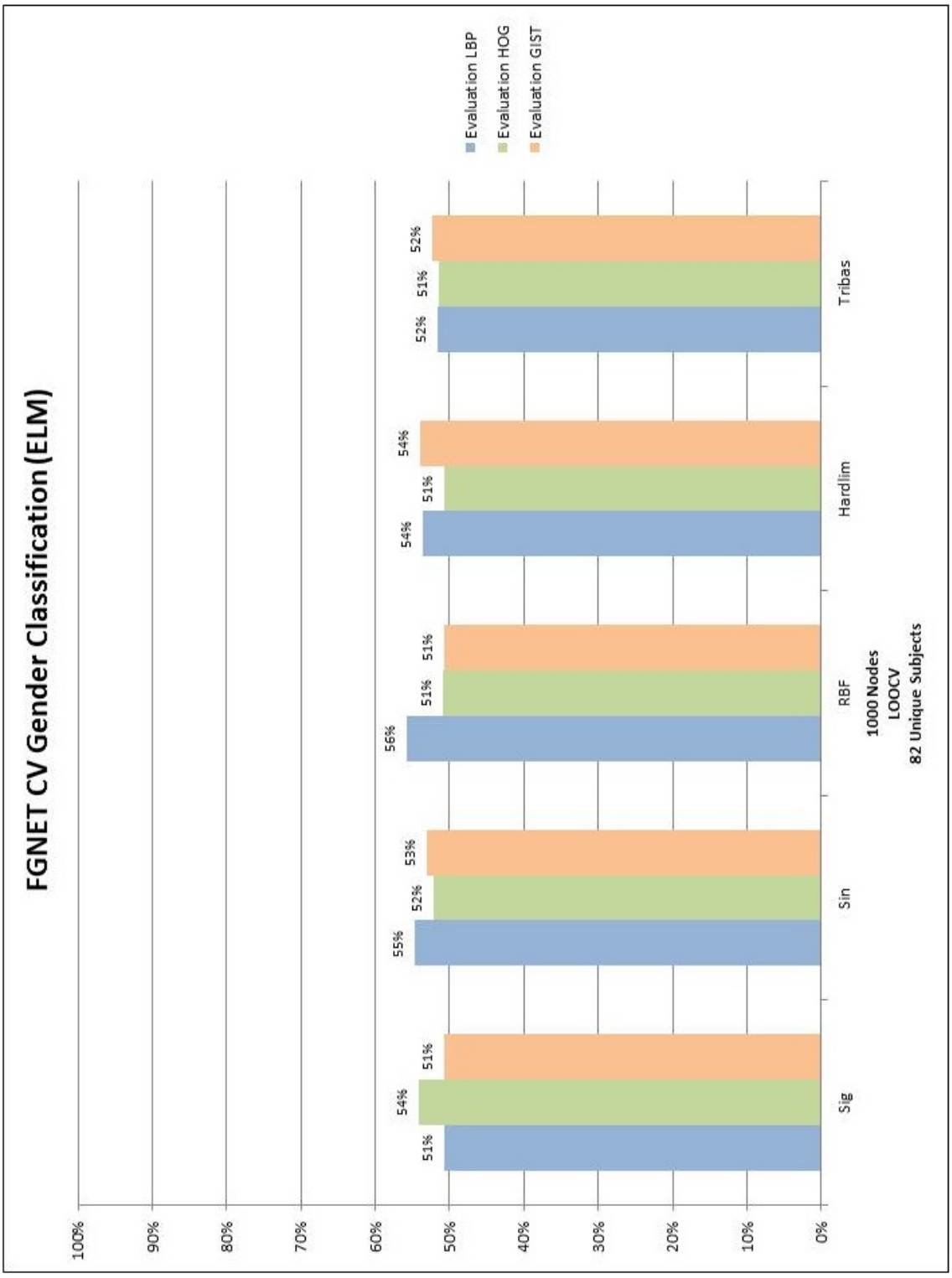

Figure 8.50: FGNET CV Gender Results

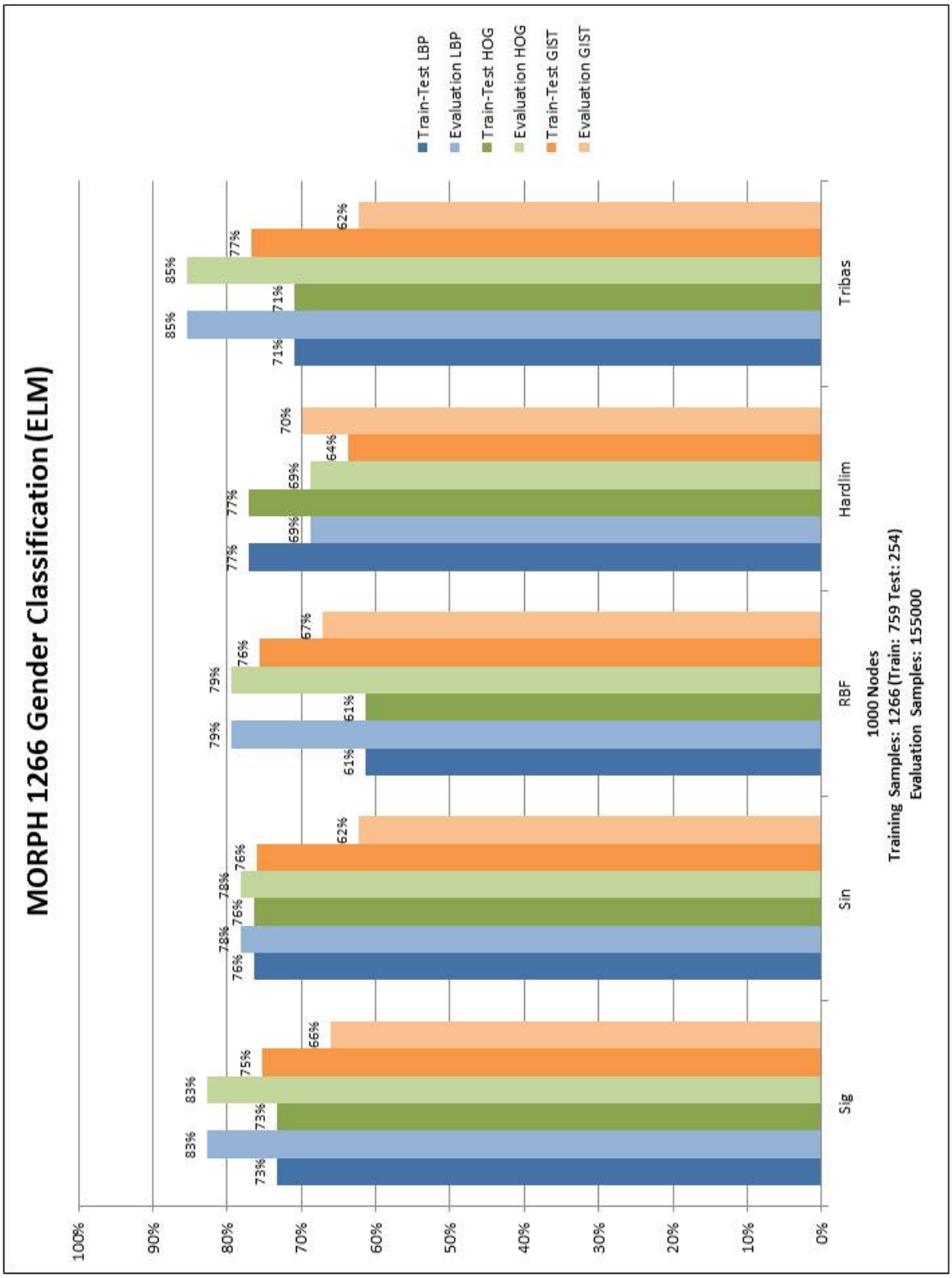

Figure 8.51: MORPH 1,266 Gender Results

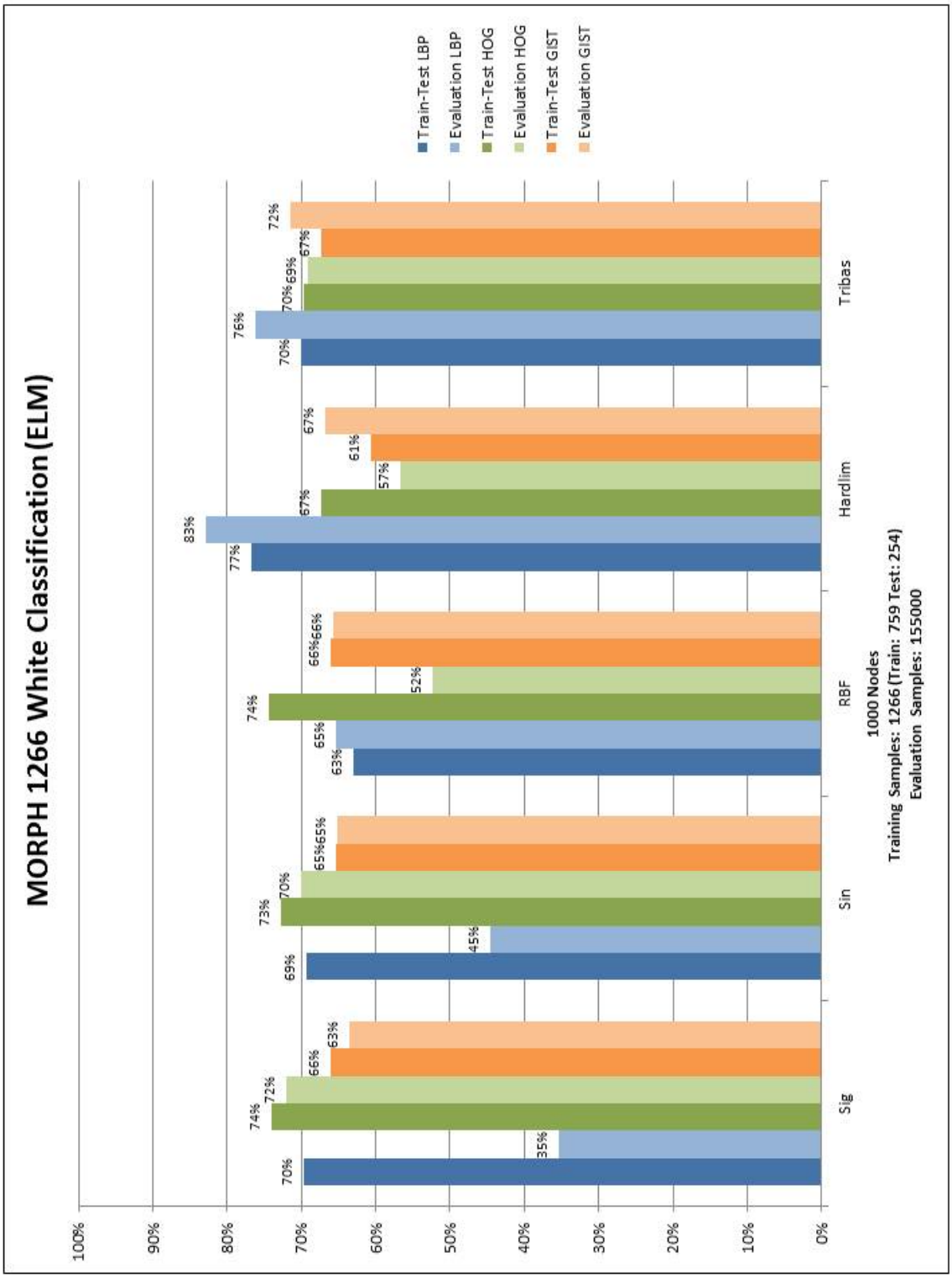

Figure 8.52: MORPH 1,266 White Results

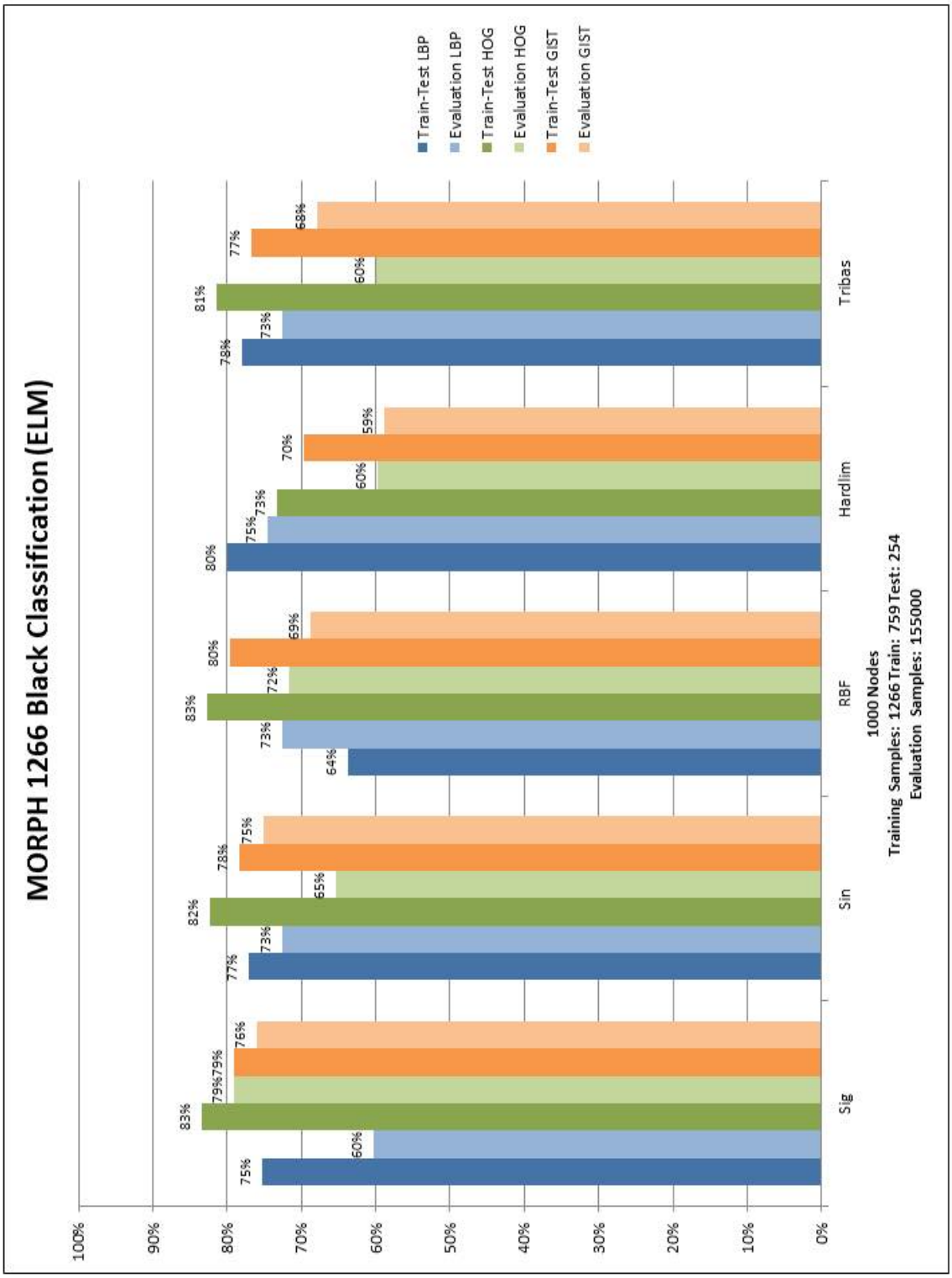

Figure 8.53: MORPH 1,266 Black Results

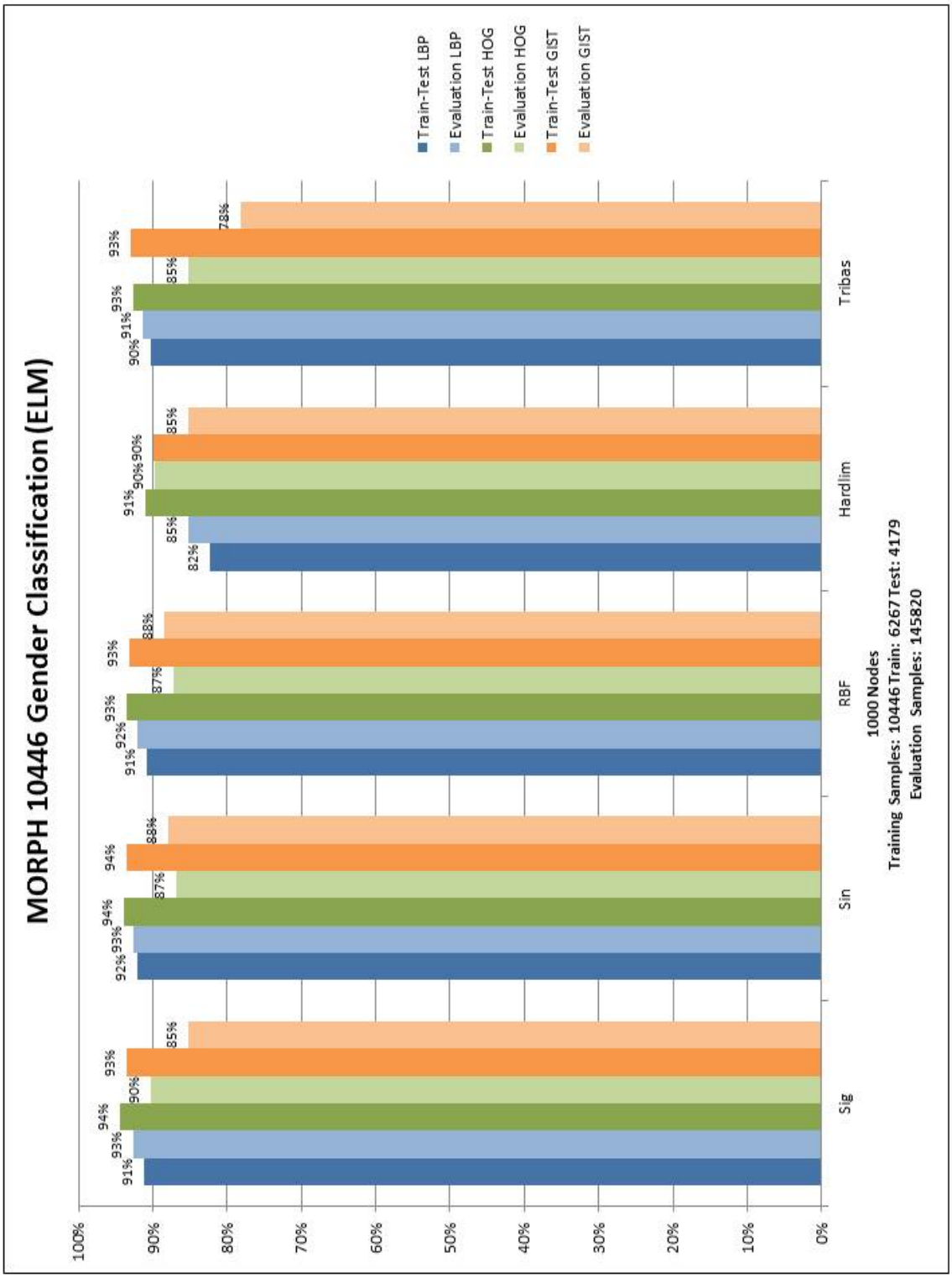

Figure 8.54: MORPH 10,446 Gender Results

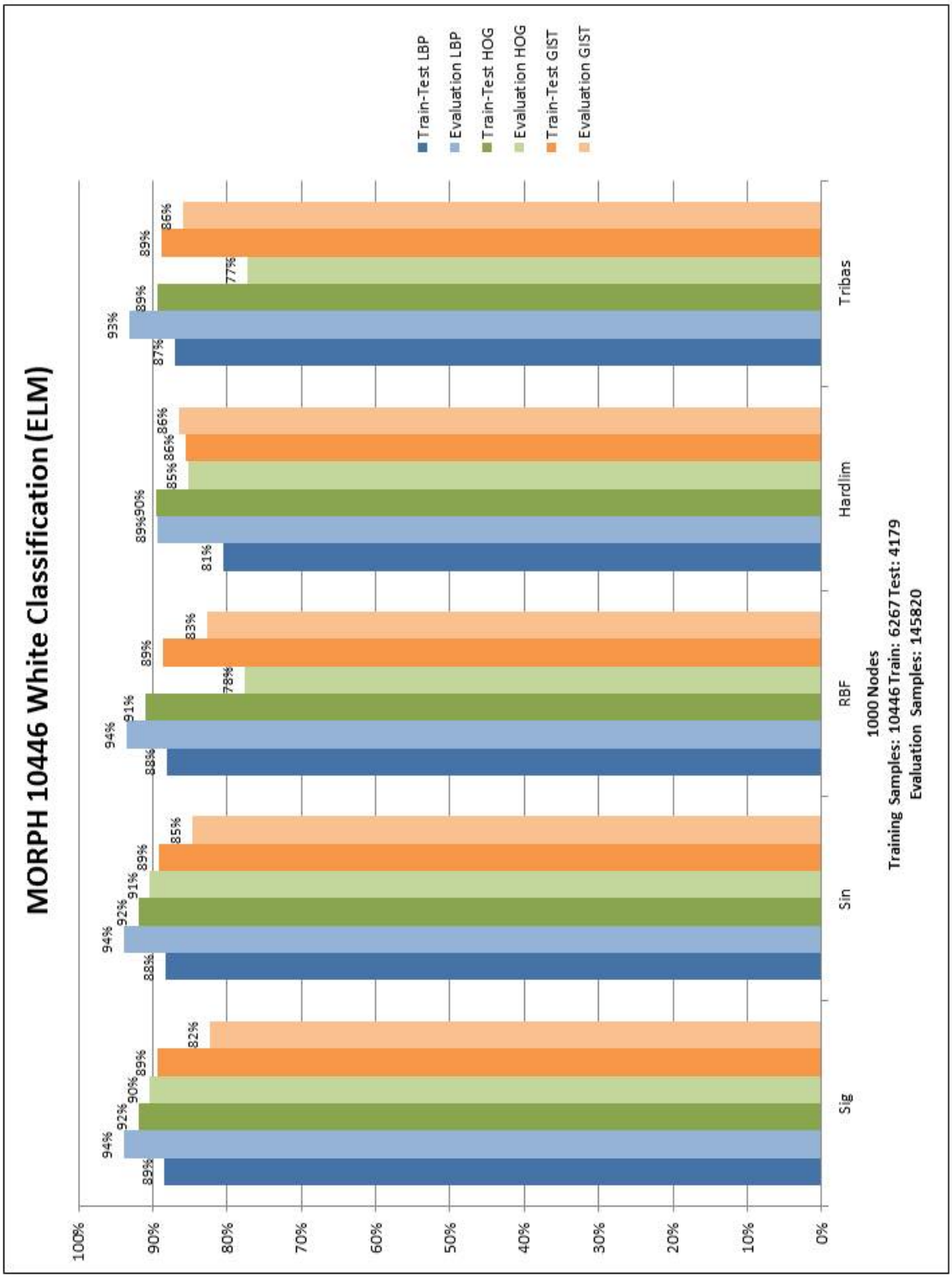

Figure 8.55: MORPH 10,446 White Results

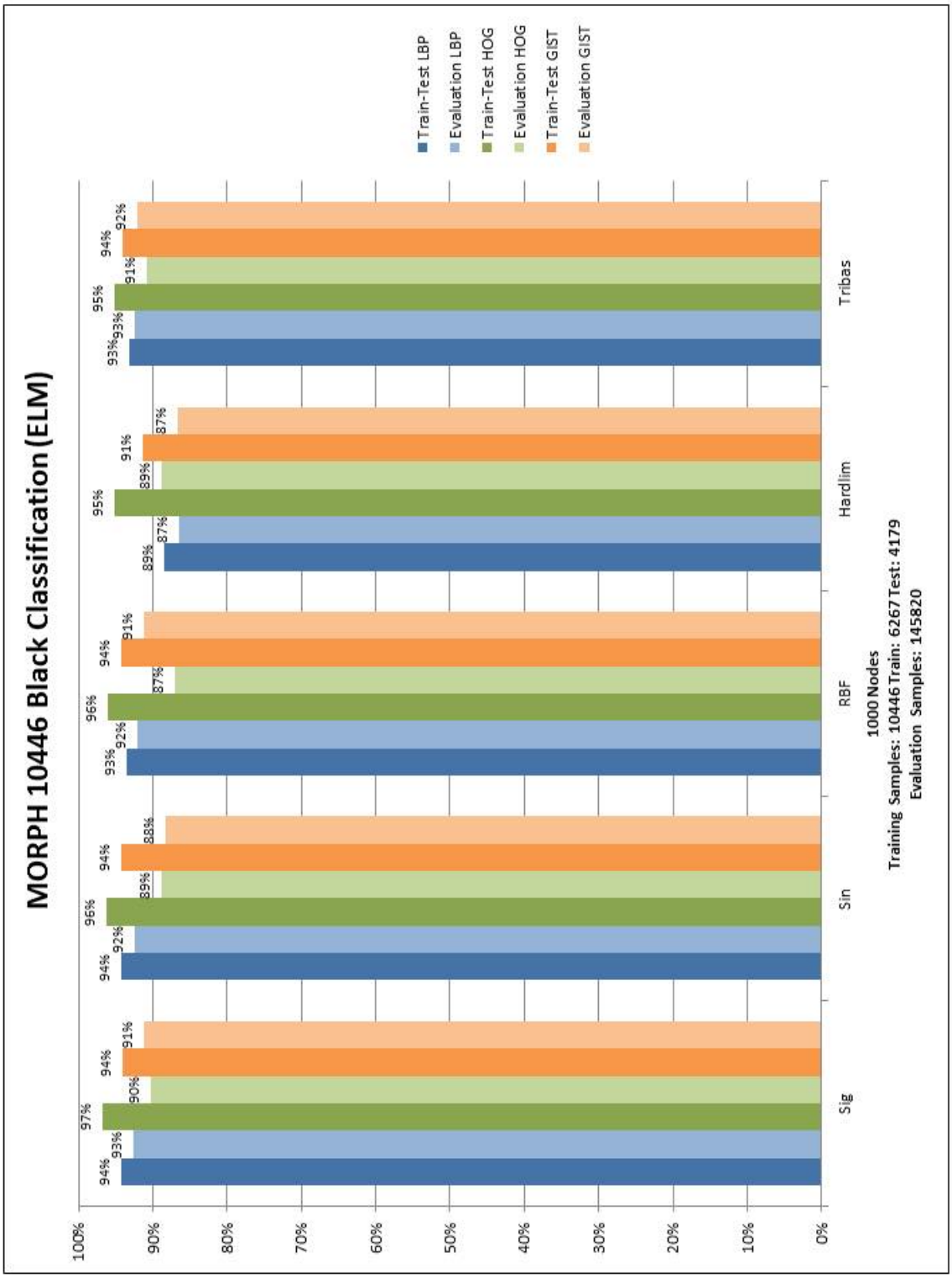

Figure 8.56: MORPH 10,446 Black Results

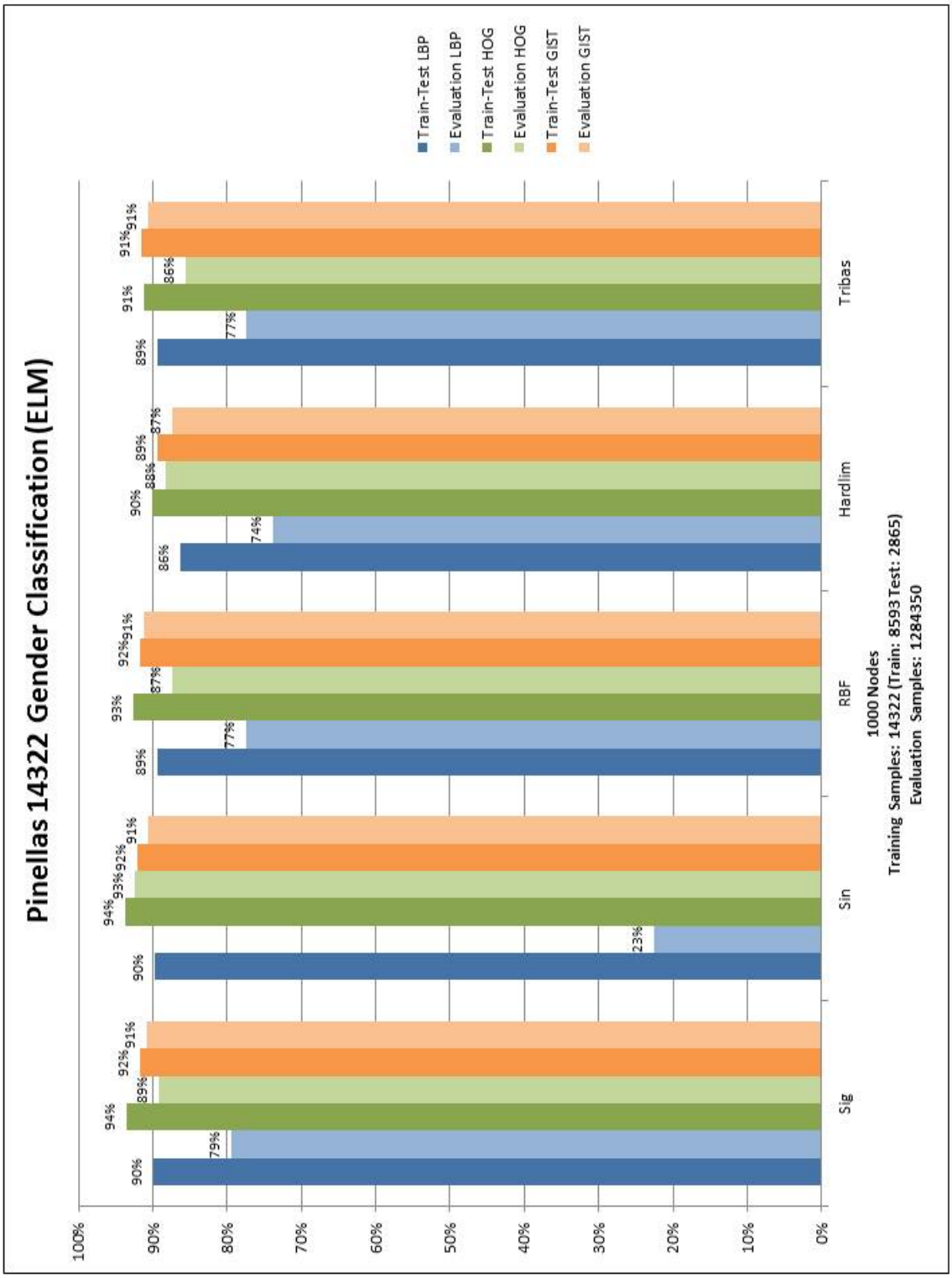

Figure 8.57: Pinellas 14,322 Gender Results

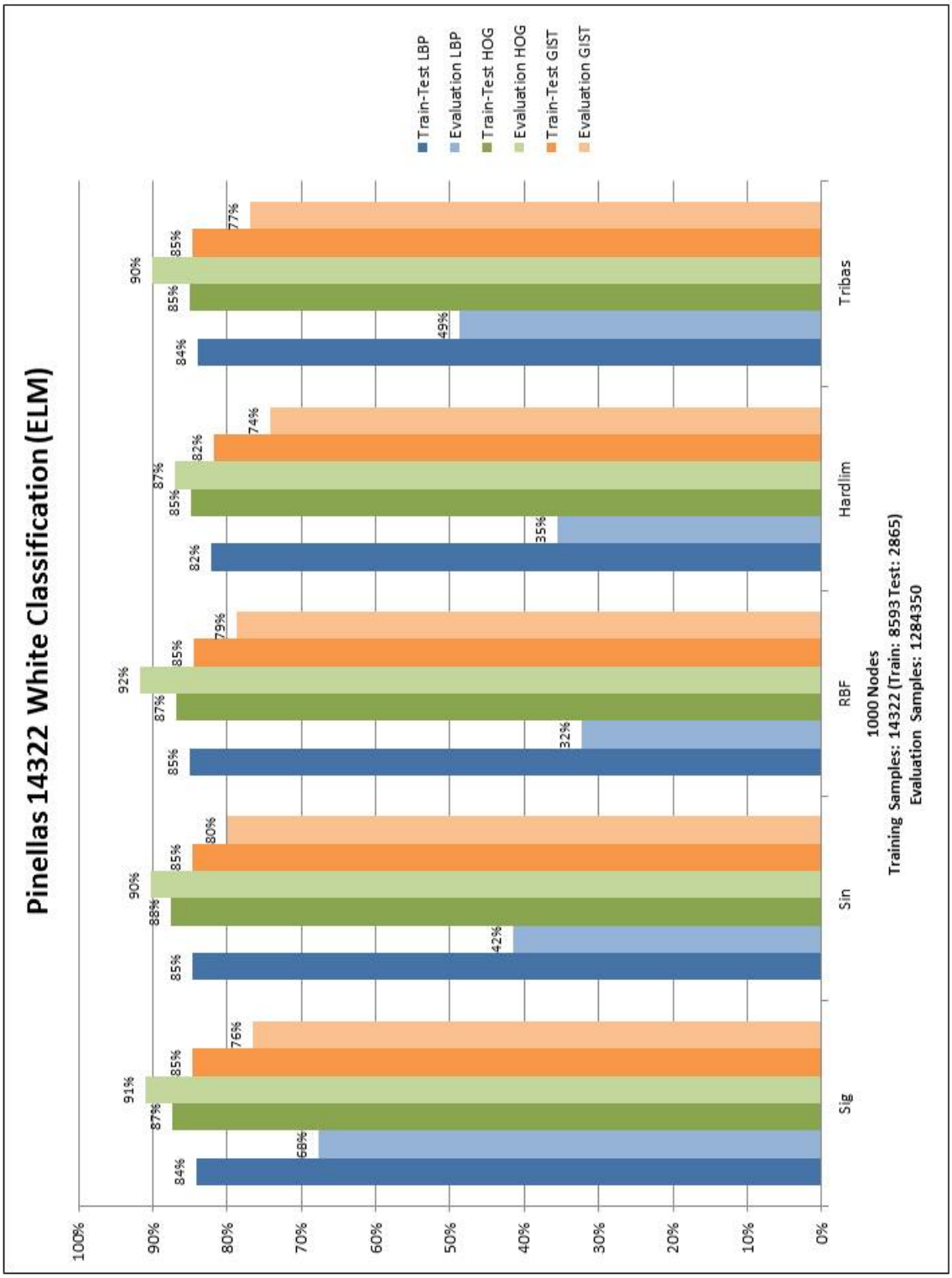

Figure 8.58: Pinellas 14,322 White Results

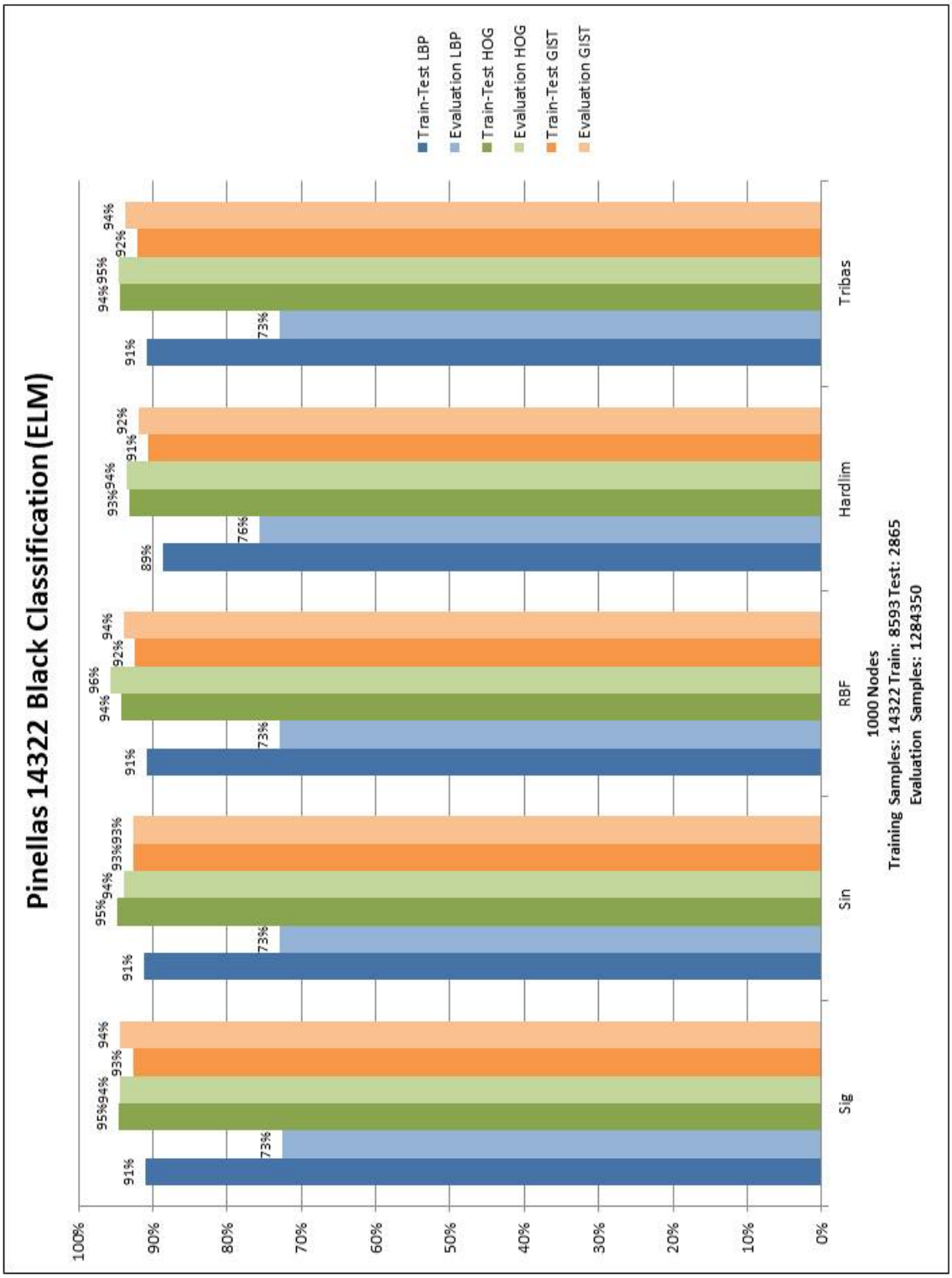

Figure 8.59: Pinellas 14,322 Black Results

Appendices G

Best Performing Results

MORPH Gender Classifiers			
	SVM	ANN	ELM
LBP	**93.3665%**	92.1880%	92.5977%
HOG	90.0007%	87.9650%	90.2345%
GIST	87.8690%	84.5070%	88.4591%

Table 8.39: Best MORPH Gender Classification Performance

MORPH White Classifiers			
	SVM	ANN	ELM
LBP	94.7890%	93.4620%	93.5400%
HOG	**95.8250%**	94.6350%	90.5000%
GIST	84.8230%	89.4420%	86.4200%

Table 8.40: Best MORPH White Classification Performance

MORPH Black Classifiers			
	SVM	ANN	ELM
LBP	94.1100%	91.9000%	92.6642%
HOG	90.3690%	—%	90.8627%
GIST	93.5330%	88.1900%	91.9894%

Table 8.41: Best MORPH Black Classification Performance

Pinellas Gender Classifiers			
	SVM	ANN	ELM
LBP	22.5540%	77.4460%	79.4020%
HOG	**94.8100%**	93.5520%	92.5190%
GIST	92.2970%	89.4320%	91.1940%

Table 8.42: Best Pinellas Gender Classification Performance

Pinellas White Classifiers			
	SVM	ANN	ELM
LBP	67.7600%	81.0590%	67.7600%
HOG	**94.4200%**	92.2680%	91.6300%
GIST	82.3000%	81.1270%	80.0400%

Table 8.43: Best Pinellas White Classification Performance

Pinellas Black Classifiers			
	SVM	ANN	ELM
LBP	72.9070%	72.9100%	72.9295%
HOG	**96.1200%**	94.9100%	95.6093%
GIST	95.1580%	94.2900%	94.4739%

Table 8.44: Best Pinellas Black Classification Performance

Appendices H

Performing Classifiers

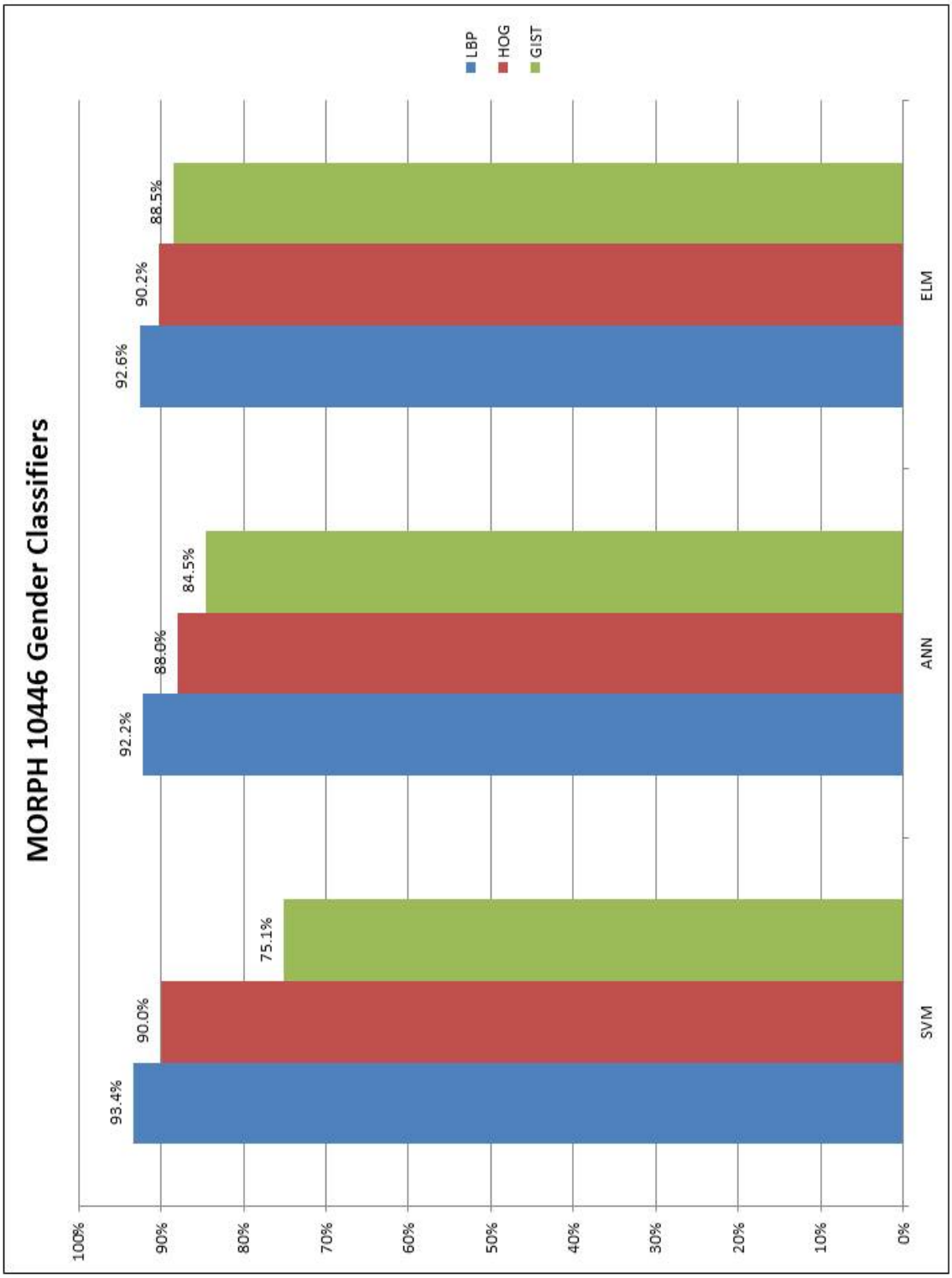

Figure 8.60: MORPH 10446 Gender Classifiers

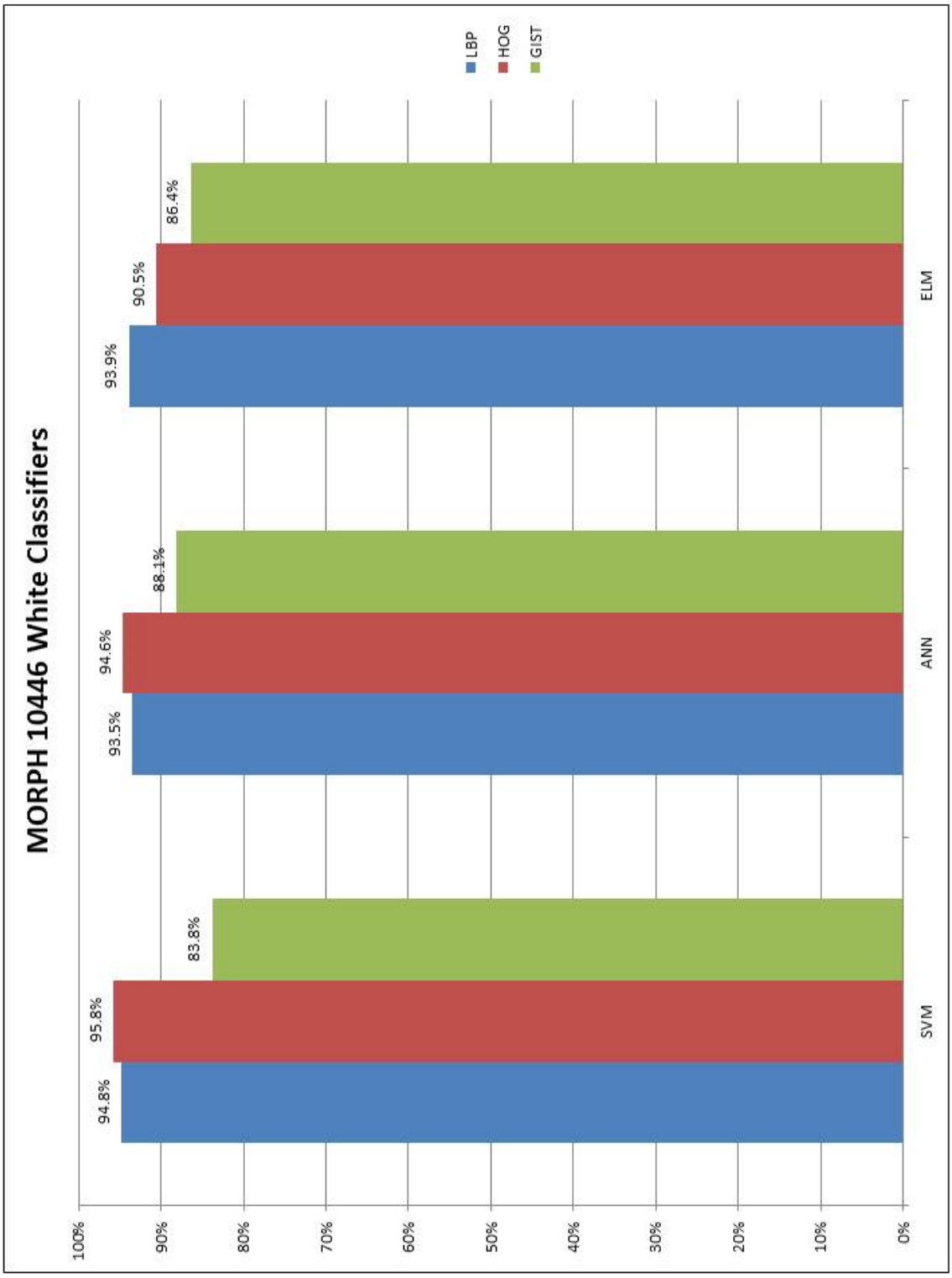

Figure 8.61: MORPH 10446 White Classifiers

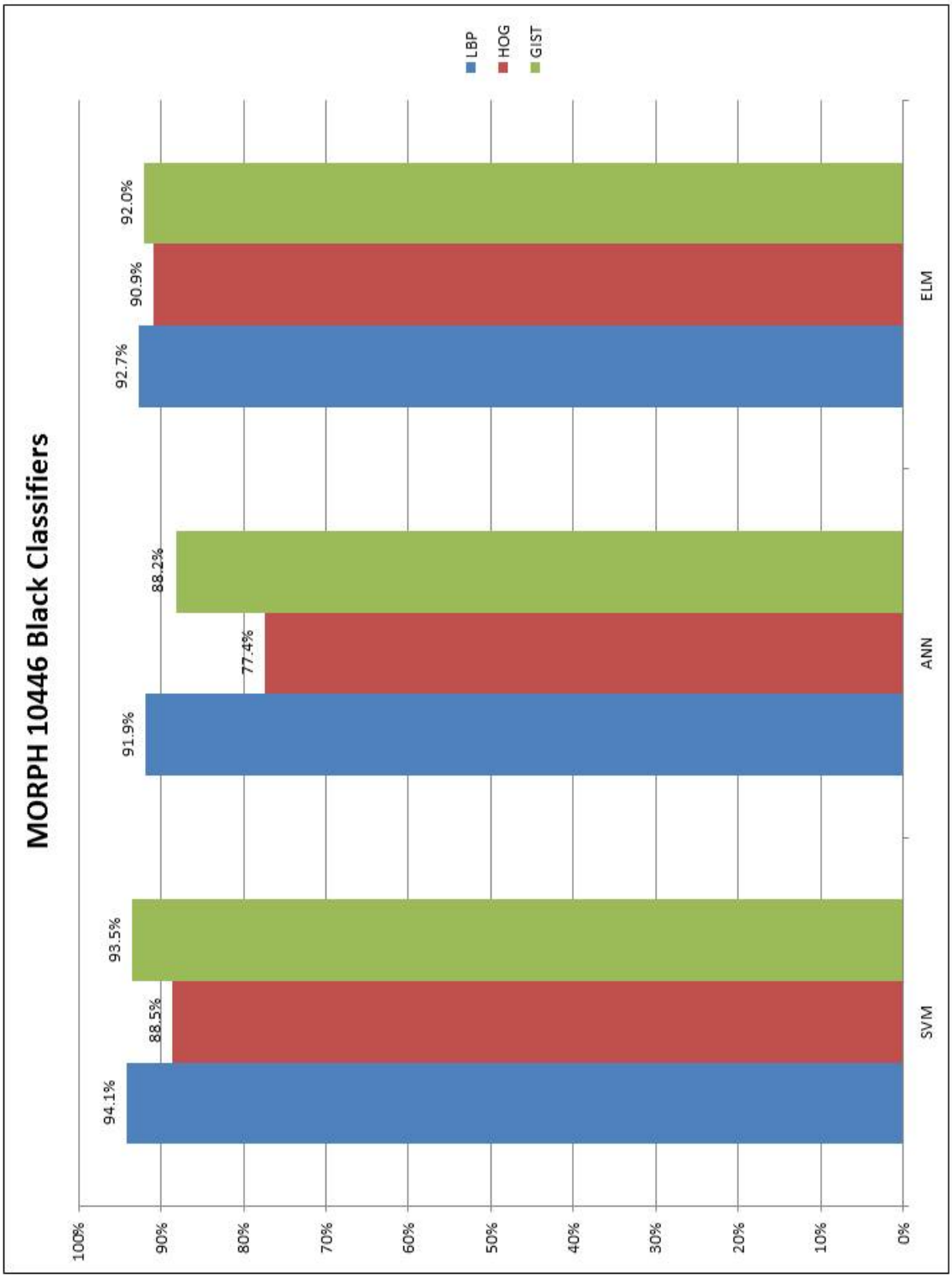

Figure 8.62: MORPH 10446 Black Classifiers

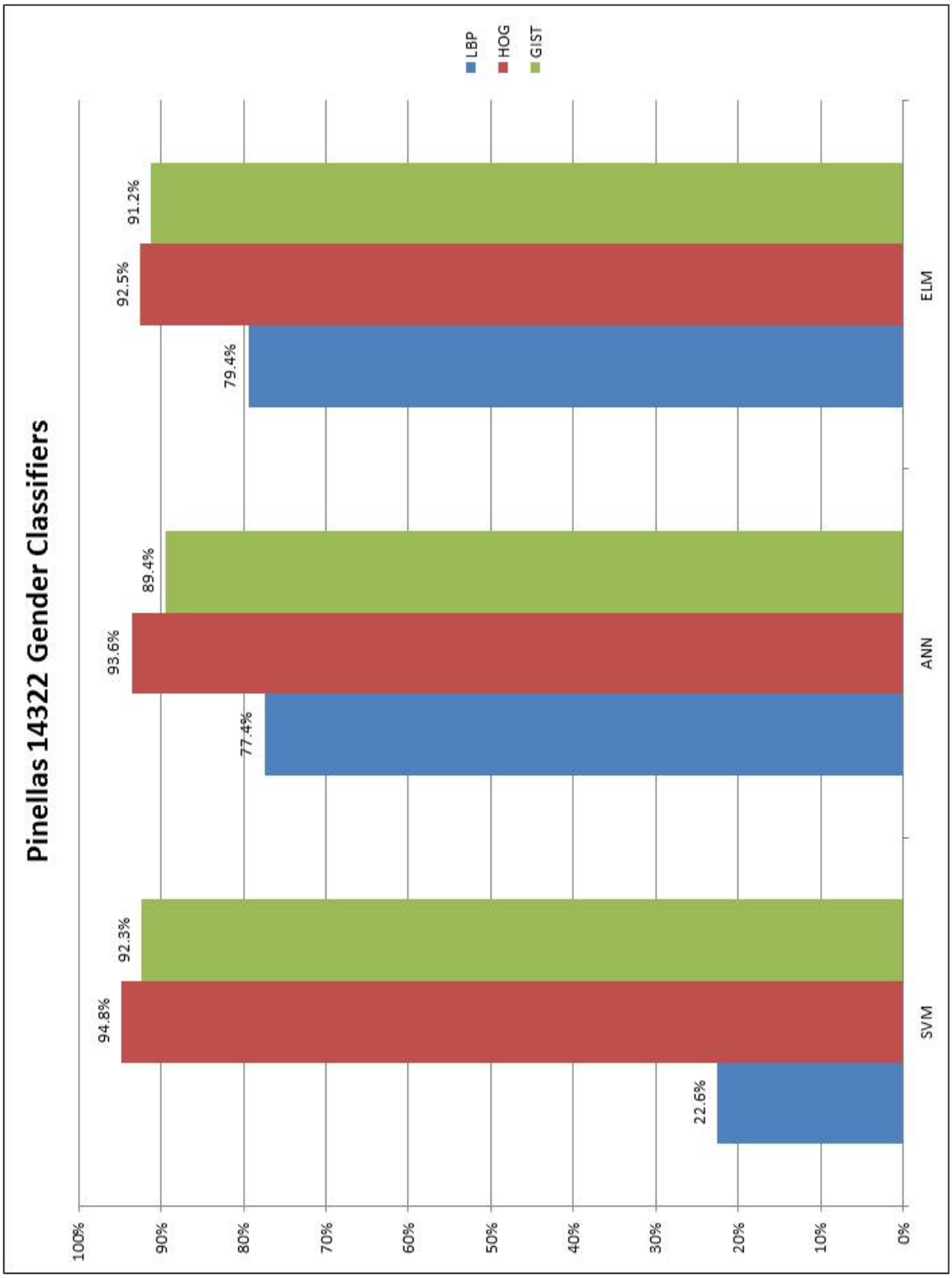

Figure 8.63: Pinellas 14322 Gender Classifiers

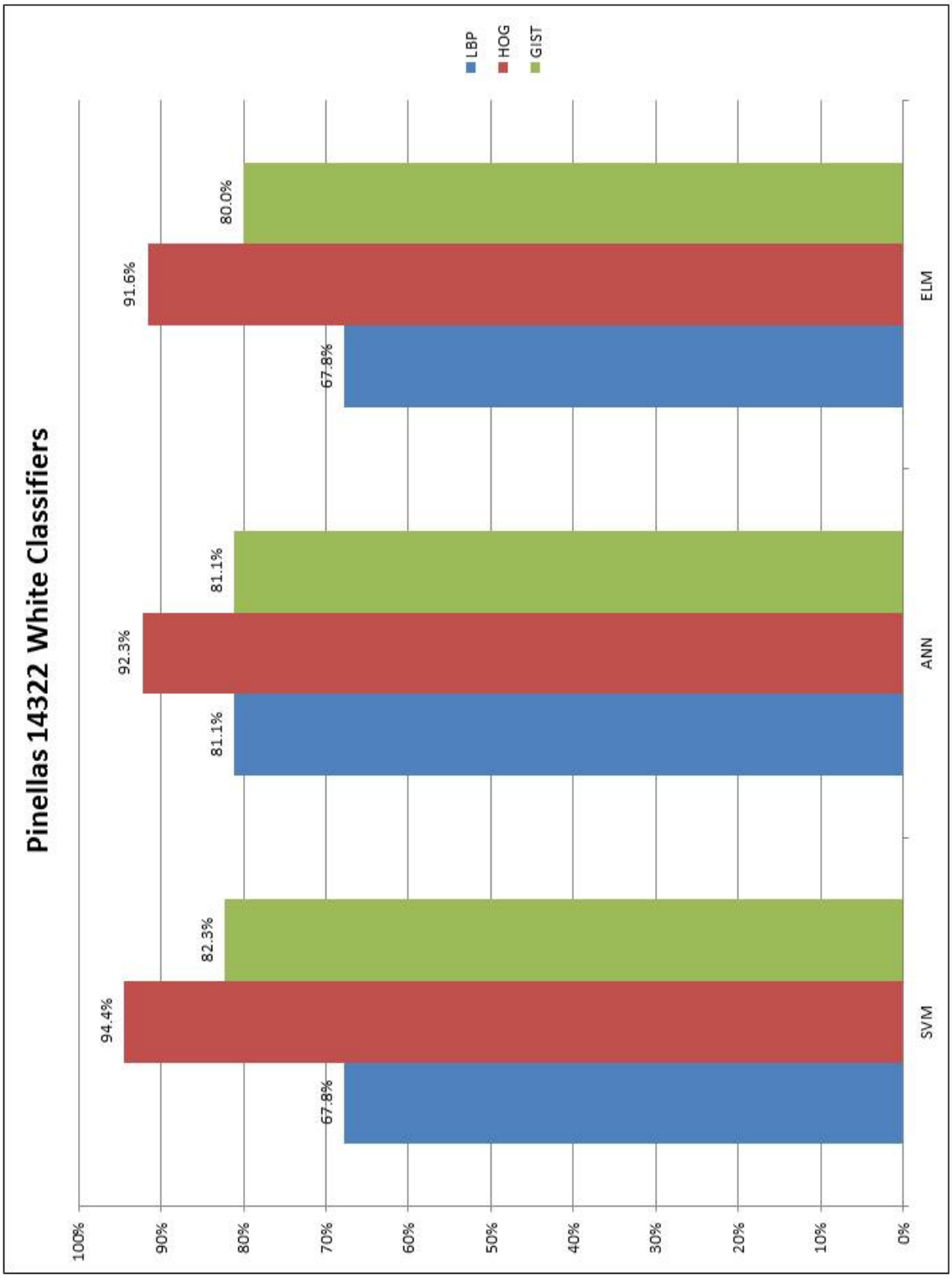

Figure 8.64: Pinellas 14322 White Classifiers

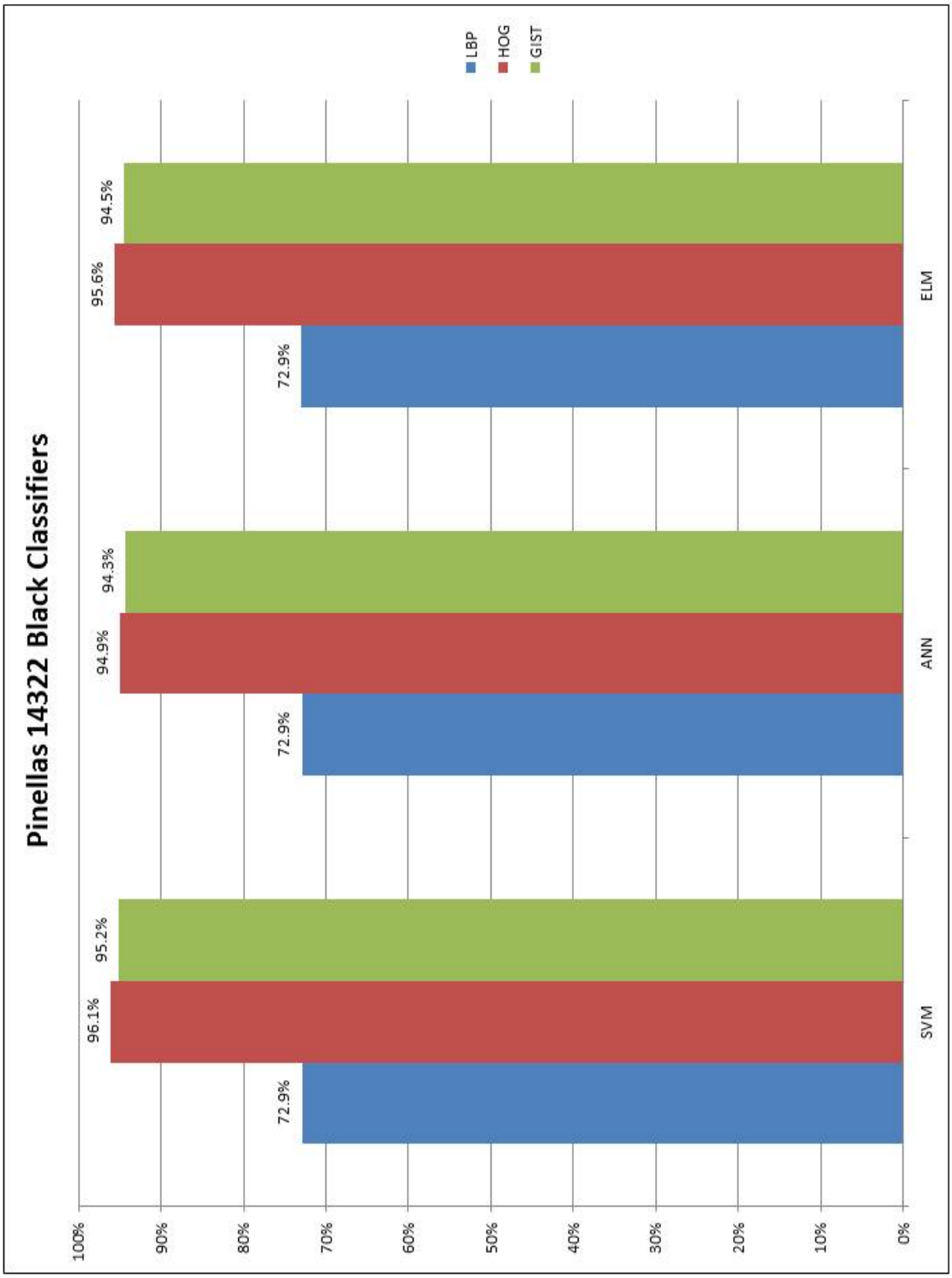

Figure 8.65: Pinellas 14322 Black Classifiers

Appendices I

Cross Dataset Evaluation Results

Gender			
	SVM-LBP	SVM-HOG	SVM-GIST
MORPH Trained	93.37%	90.00%	75.09%
Pinellas Tested	77.97%	91.40%	25.22%
Gender			
	ANN-LBP	ANN-HOG	ANN-GIST
MORPH Trained	92.19%	87.96%	84.51%
Pinellas Tested	77.36%	88.64%	76.15%
Gender			
	ELM-LBP	ELM-HOG	ELM-GIST
MORPH Trained	92.60%	90.24%	88.46%
Pinellas Tested	77.70%	86.56%	77.94%

Table 8.45: Cross Dataset Experiment (Gender - MORPH Trained/Pinellas Tested)

White			
	SVM-LBP	SVM-HOG	SVM-GIST
MORPH Trained	94.79%	95.83%	83.81%
Pinellas Tested	69.27%	82.97%	32.69%
White			
	ANN-LBP	ANN-HOG	ANN-GIST
MORPH Trained	93.46%	94.64%	88.10%
Pinellas Tested	63.32%	65.64%	35.32%
White			
	ELM-LBP	ELM-HOG	ELM-GIST
MORPH Trained	93.86%	90.50%	86.42%
Pinellas Tested	39.80%	78.92%	37.81%

Table 8.46: Cross Dataset Experiment (White - MORPH Trained/Pinellas Tested)

Black			
	SVM-LBP	SVM-HOG	SVM-GIST
MORPH Trained	94.11%	88.55%	93.53%
Pinellas Tested	72.84%	96.13%	28.08%
Black			
	ANN-LBP	ANN-HOG	ANN-GIST
MORPH Trained	91.90%	77.38%	88.19%
Pinellas Tested	72.84%	95.45%	28.50%
Black			
	ELM-LBP	ELM-HOG	ELM-GIST
MORPH Trained	92.66%	90.86%	91.99%
Pinellas Tested	77.20%	91.41%	60.16%

Table 8.47: Cross Dataset Experiment (Black - MORPH Trained/Pinellas Tested)

Gender			
	SVM-LBP	SVM-HOG	SVM-GIST
Pinellas Trained	22.55%	94.81%	92.30%
MORPH Tested	16.74%	93.58%	84.99%
Gender			
	ANN-LBP	ANN-HOG	ANN-GIST
Pinellas Trained	77.45%	93.55%	89.43%
MORPH Tested	83.32%	92.51%	83.65%
Gender			
	ELM-LBP	ELM-HOG	ELM-GIST
Pinellas Trained	79.40%	92.52%	91.19%
MORPH Tested	83.33%	91.31%	71.46%

Table 8.48: Cross Dataset Experiment (Gender - Pinellas Trained/MORPH Tested)

White			
	SVM-LBP	SVM-HOG	SVM-GIST
Pinellas Trained	67.76%	94.42%	82.31%
MORPH Tested	21.53%	79.43%	84.29%
White			
	ANN-LBP	ANN-HOG	ANN-GIST
Pinellas Trained	81.06%	92.27%	81.13%
MORPH Tested	86.51%	59.76%	80.68%
White			
	ELM-LBP	ELM-HOG	ELM-GIST
Pinellas Trained	67.76%	91.63%	80.04%
MORPH Tested	20.59%	76.42%	77.29%

Table 8.49: Cross Dataset Experiment (White - Pinellas Trained/MORPH Tested)

Black			
	SVM-LBP	SVM-HOG	SVM-GIST
Pinellas Trained	72.91%	96.12%	95.16%
MORPH Tested	26.06%	69.58%	72.57%
Black			
	ANN-LBP	ANN-HOG	ANN-GIST
Pinellas Trained	72.91%	94.91%	94.29%
MORPH Tested	26.58%	71.60%	68.25%
Black			
	ELM-LBP	ELM-HOG	ELM-GIST
Pinellas Trained	75.61%	94.67%	94.47%
MORPH Tested	52.93%	63.01%	70.58%

Table 8.50: Cross Dataset Experiment Black

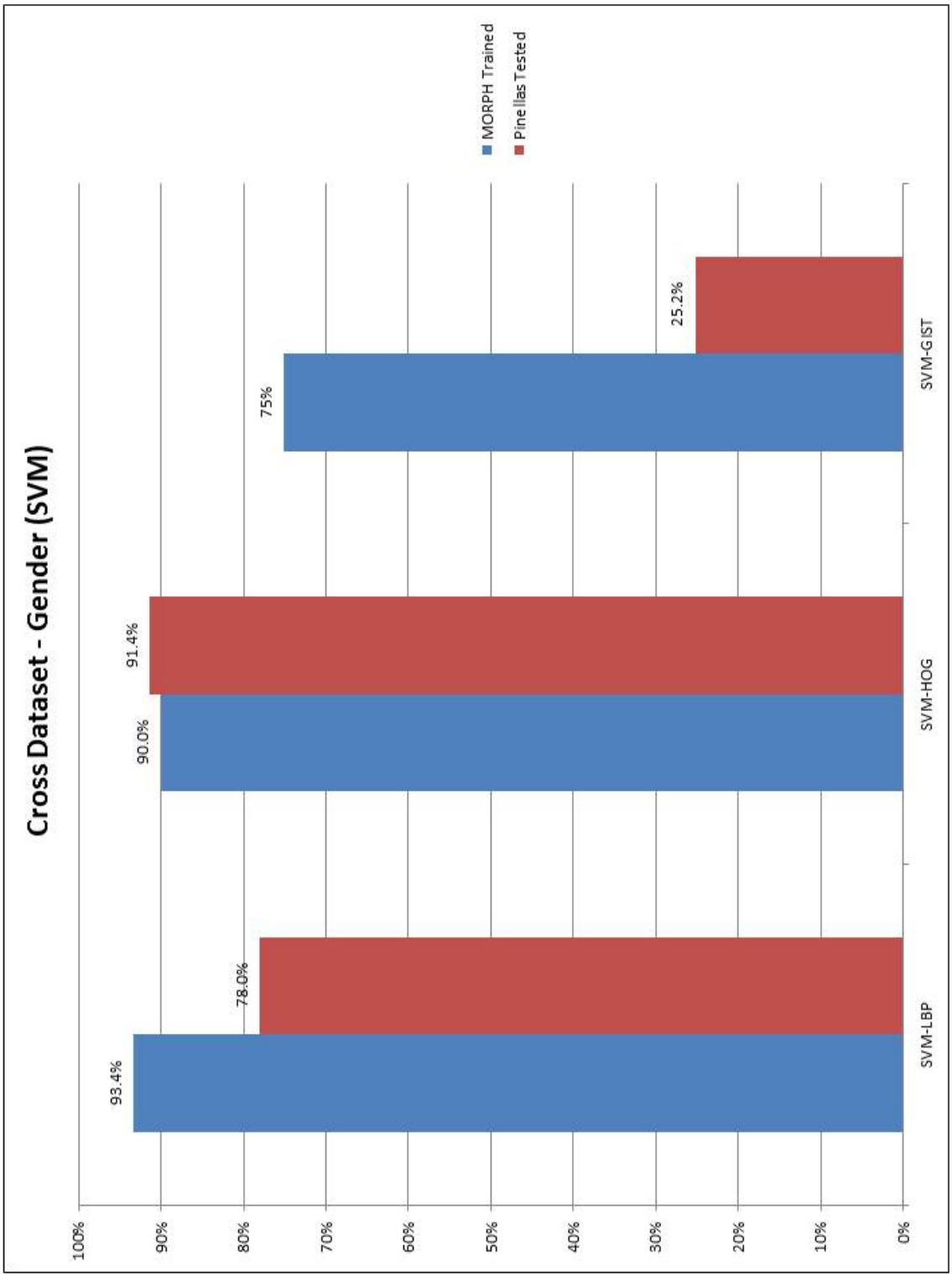

Figure 8.66: Cross Dataset - SVM Gender (MORPH Trained - Pinellas Tested)

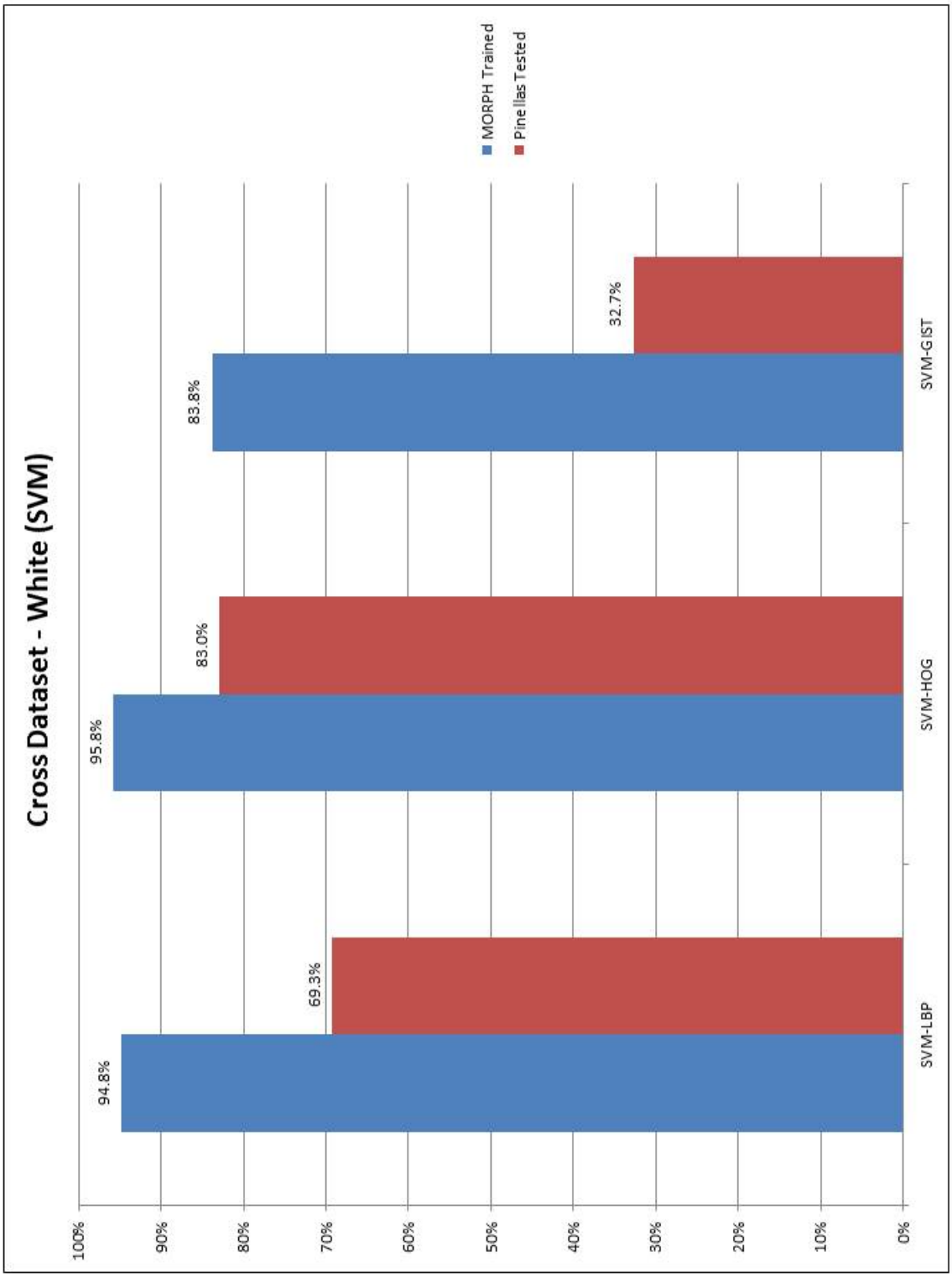

Figure 8.67: Cross Dataset - SVM White (MORPH Trained - Pinellas Tested)

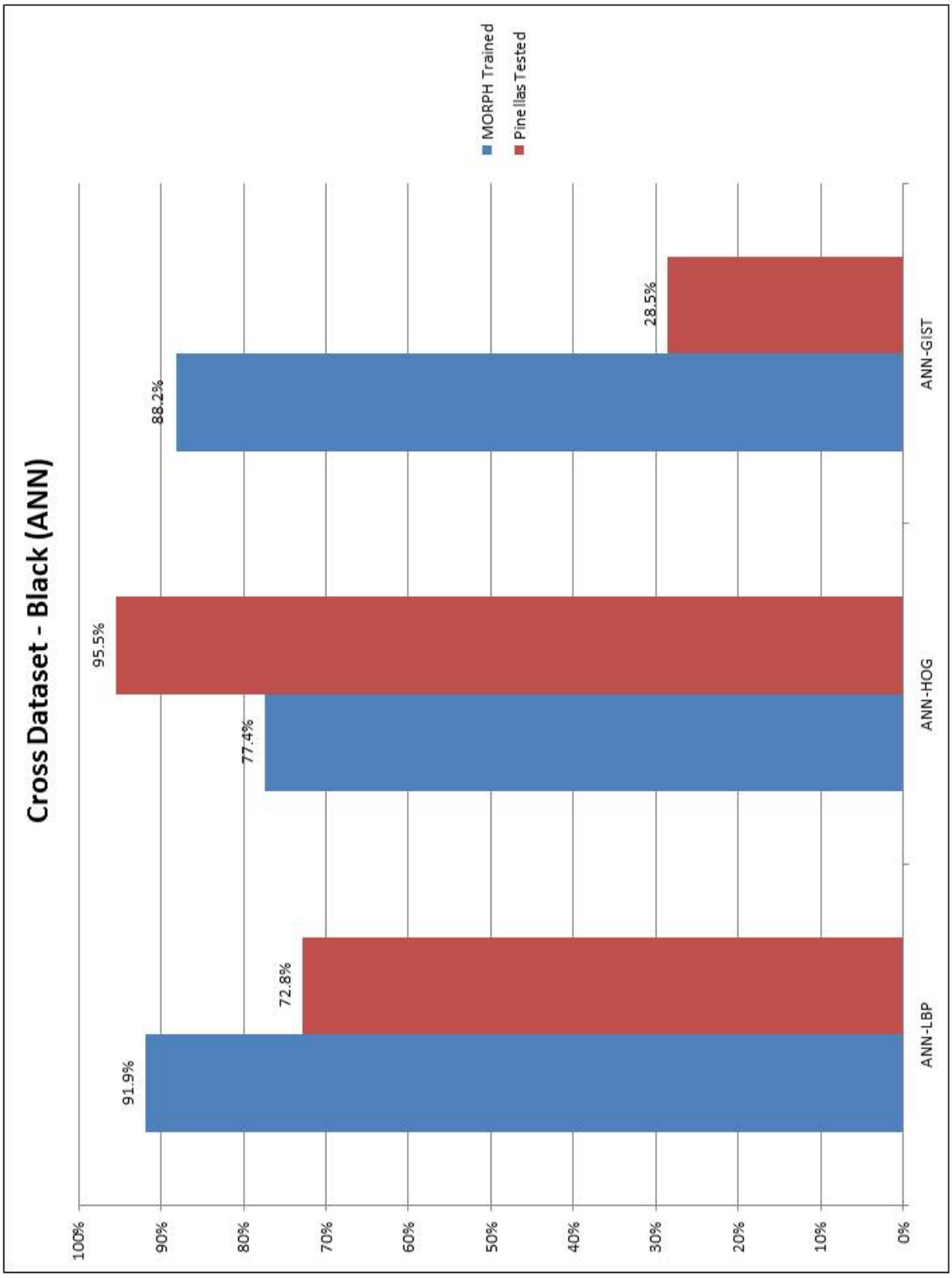

Figure 8.68: Cross Dataset - SVM Black (MORPH Trained - Pinellas Tested)

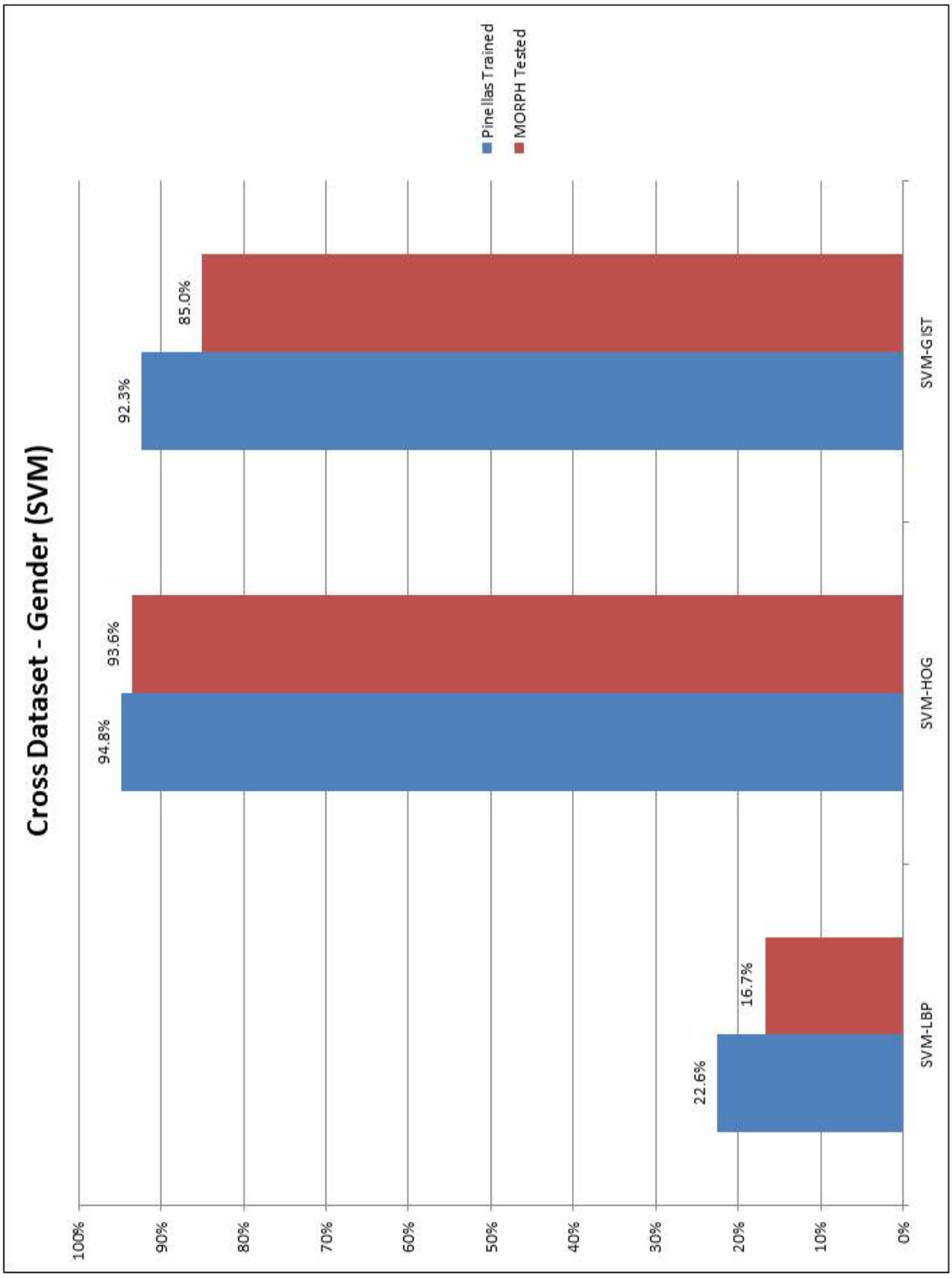

Figure 8.69: Cross Dataset - SVM Gender (Pinellas Trained - MORPH Tested)

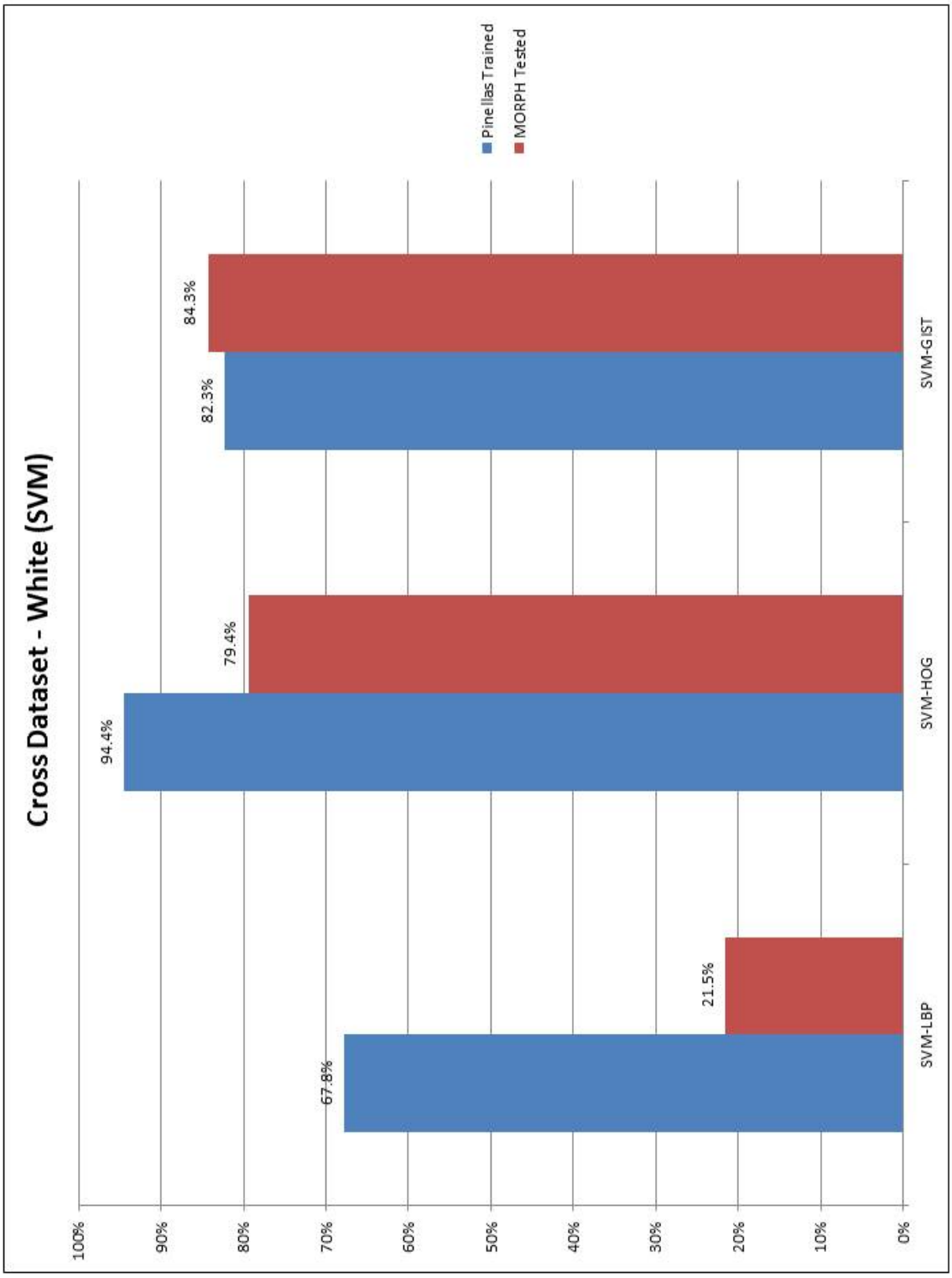

Figure 8.70: Cross Dataset - SVM White (Pinellas Trained - MORPH Tested)

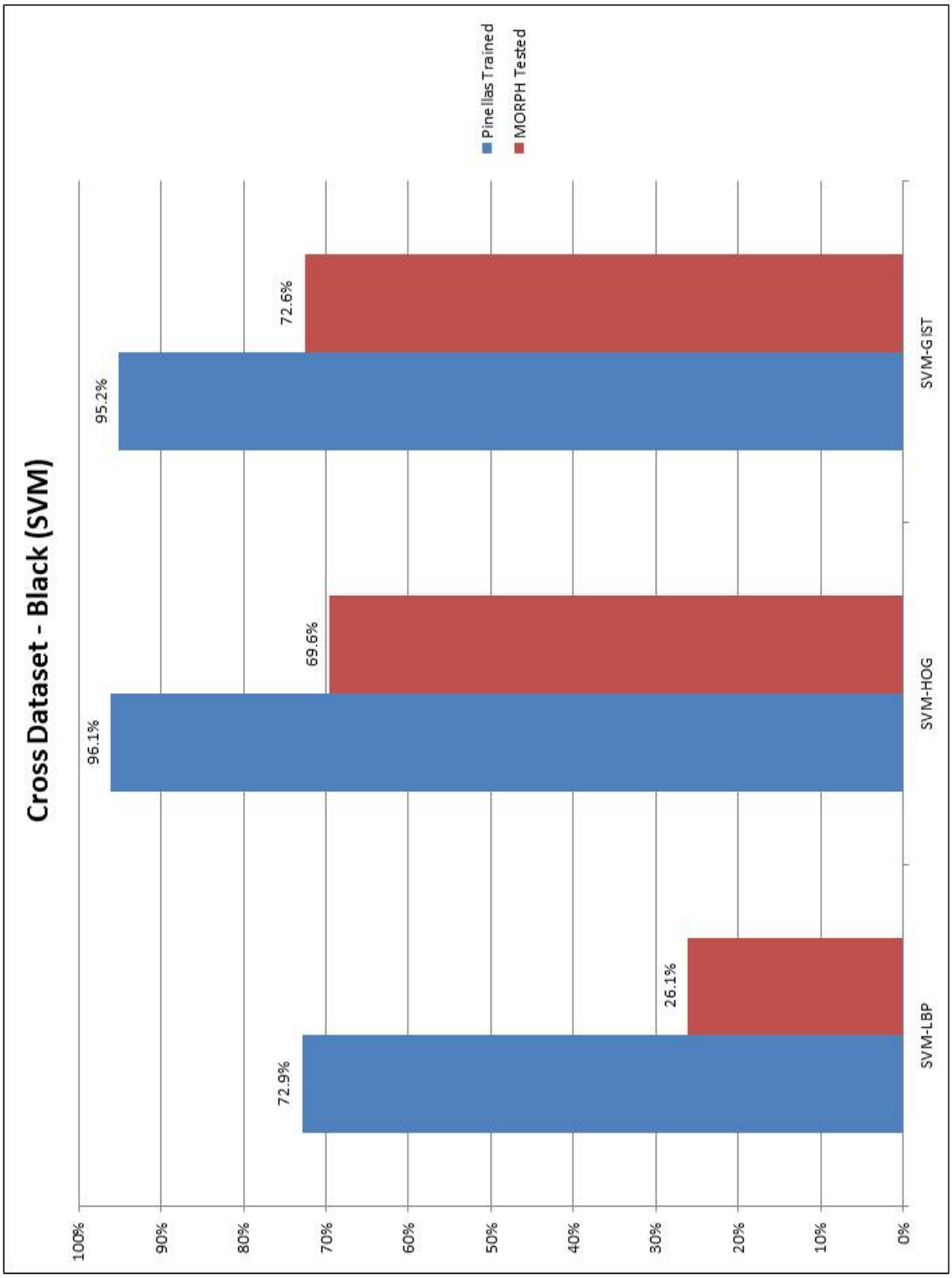

Figure 8.71: Cross Dataset - SVM Black (Pinellas Trained - MORPH Tested)

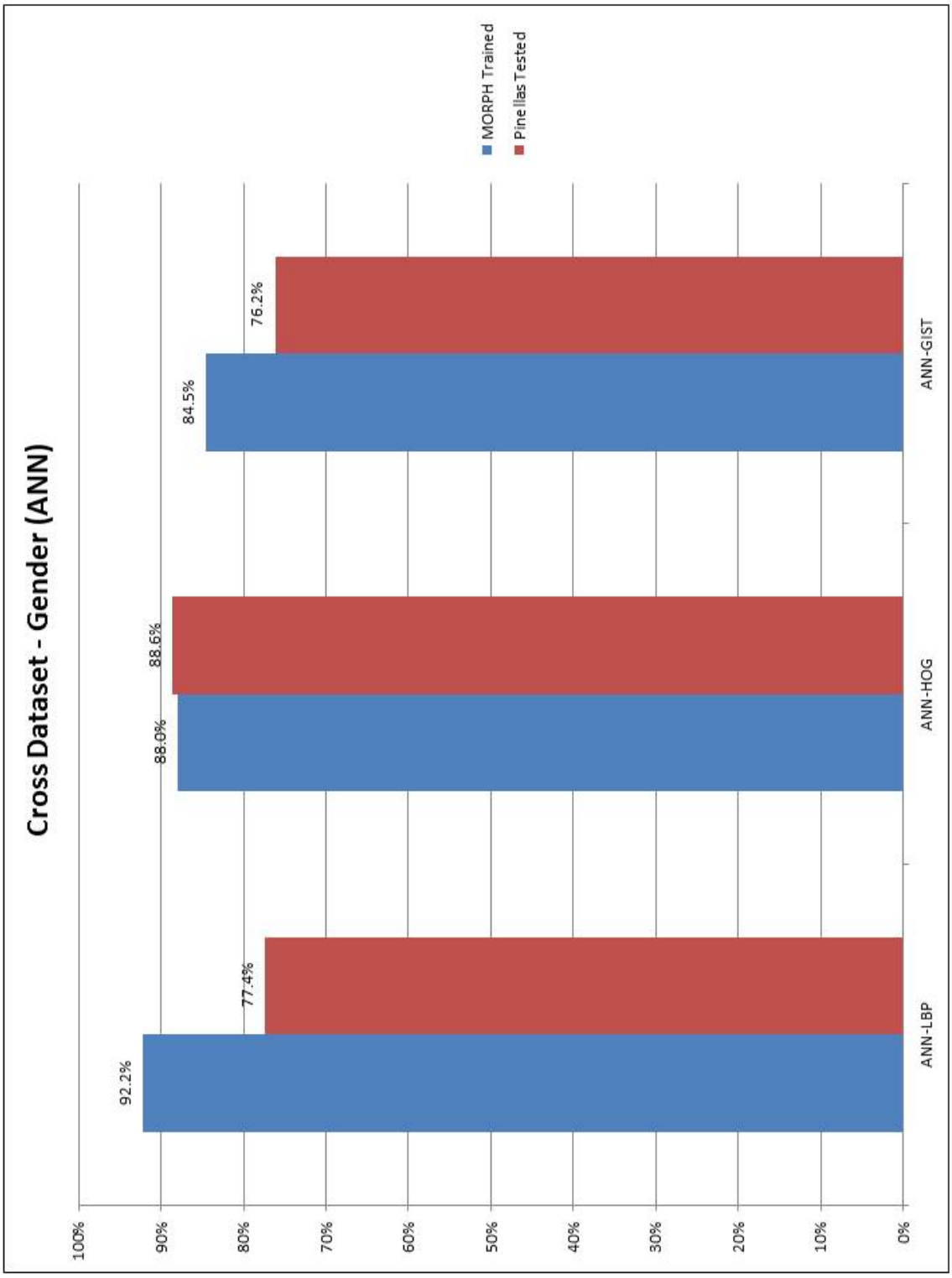

Figure 8.72: Cross Dataset - ANN Gender (MORPH Trained - Pinellas Tested)

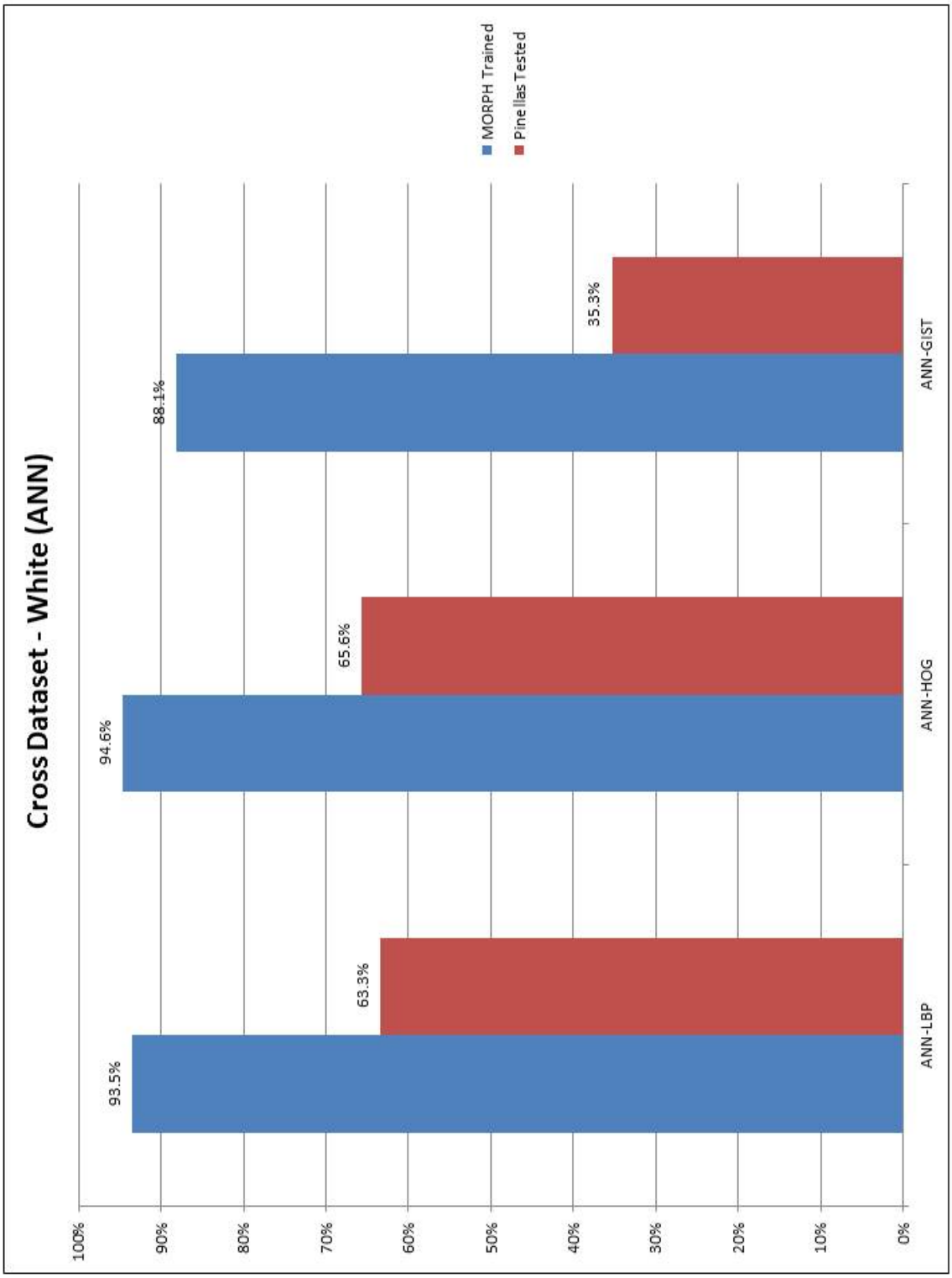

Figure 8.73: Cross Dataset - ANN White (MORPH Trained - Pinellas Tested)

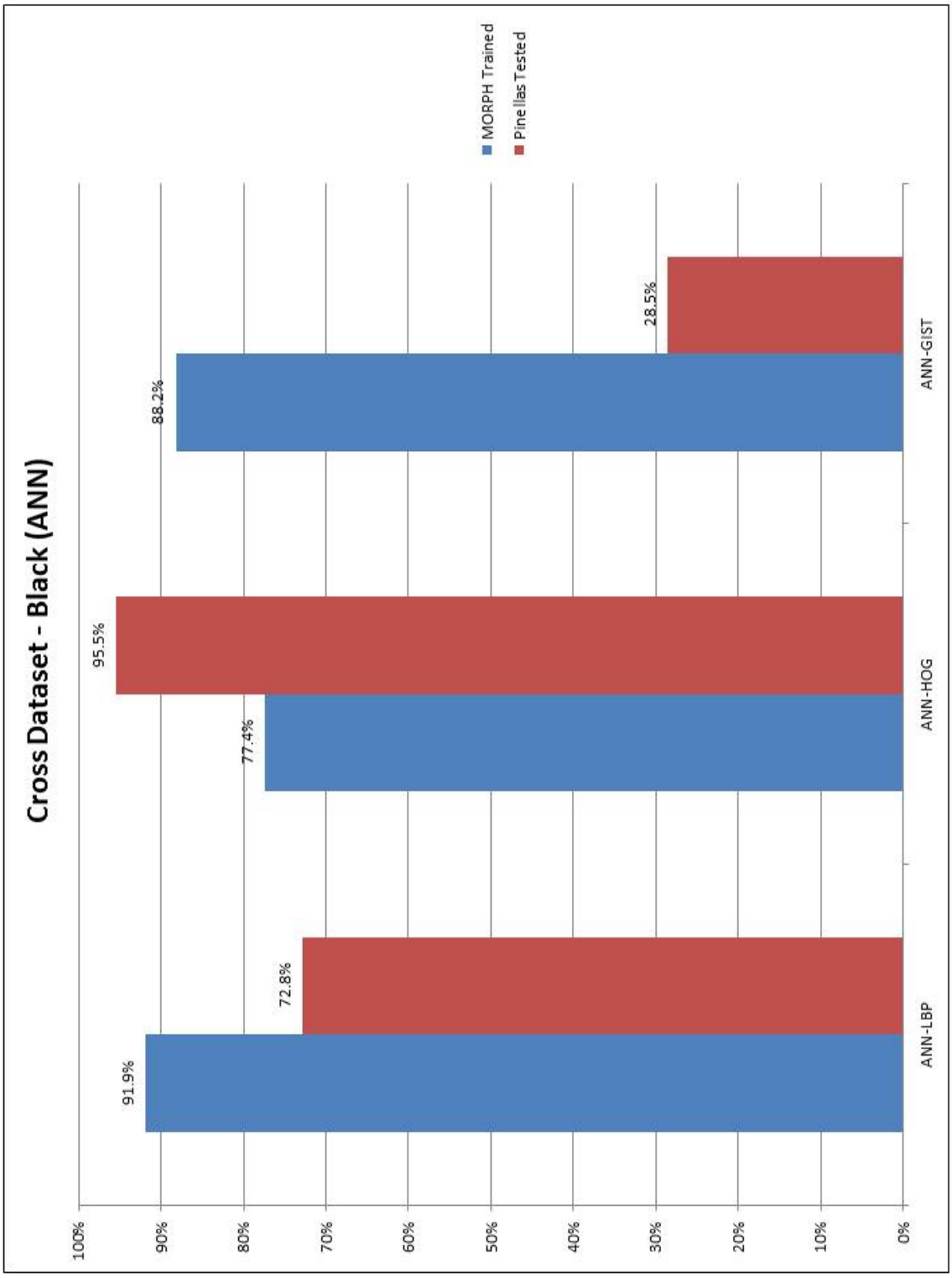

Figure 8.74: Cross Dataset - ANN Black (MORPH Trained - Pinellas Tested)

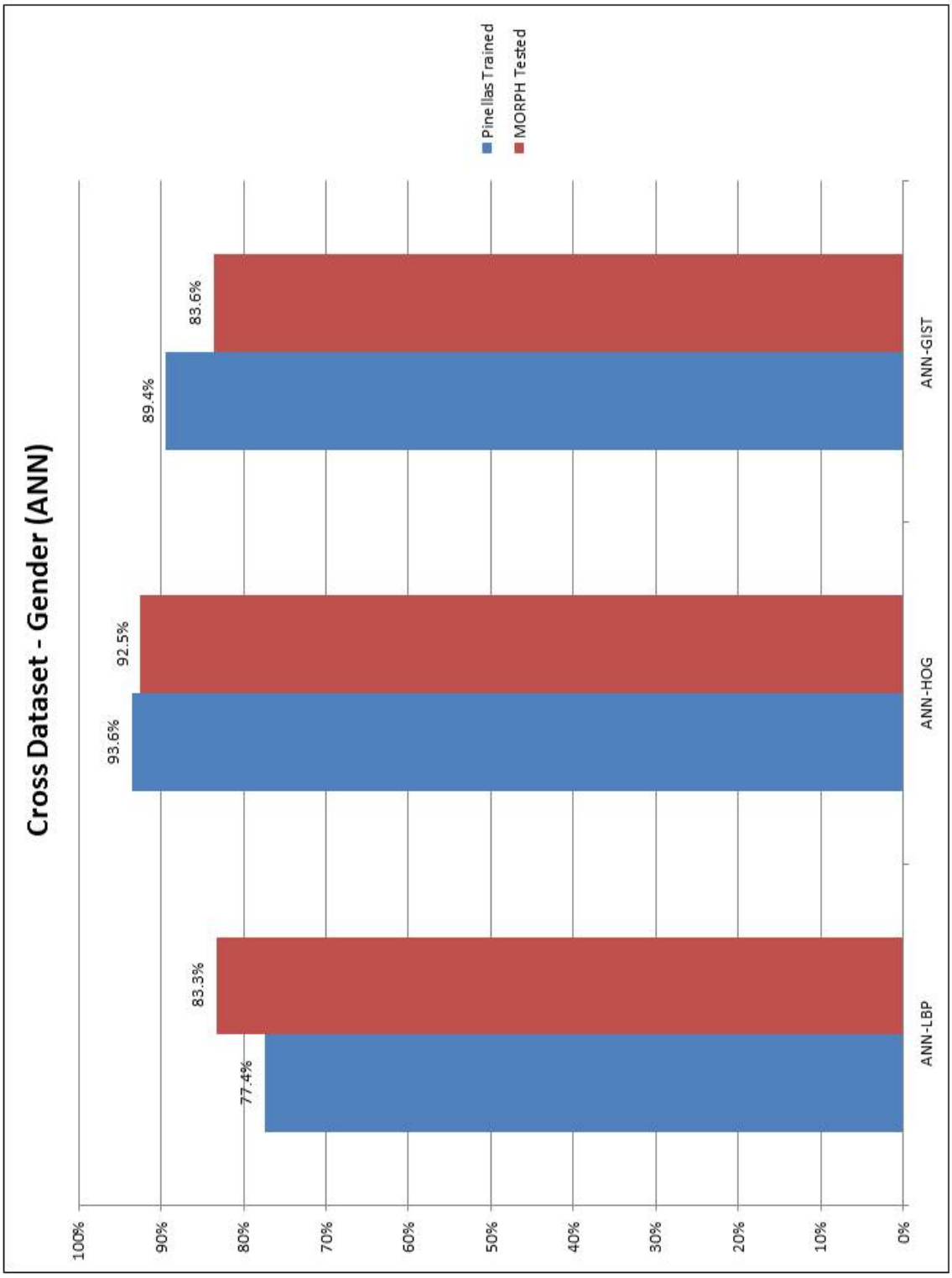

Figure 8.75: Cross Dataset - ANN Gender (Pinellas Trained - MORPH Tested)

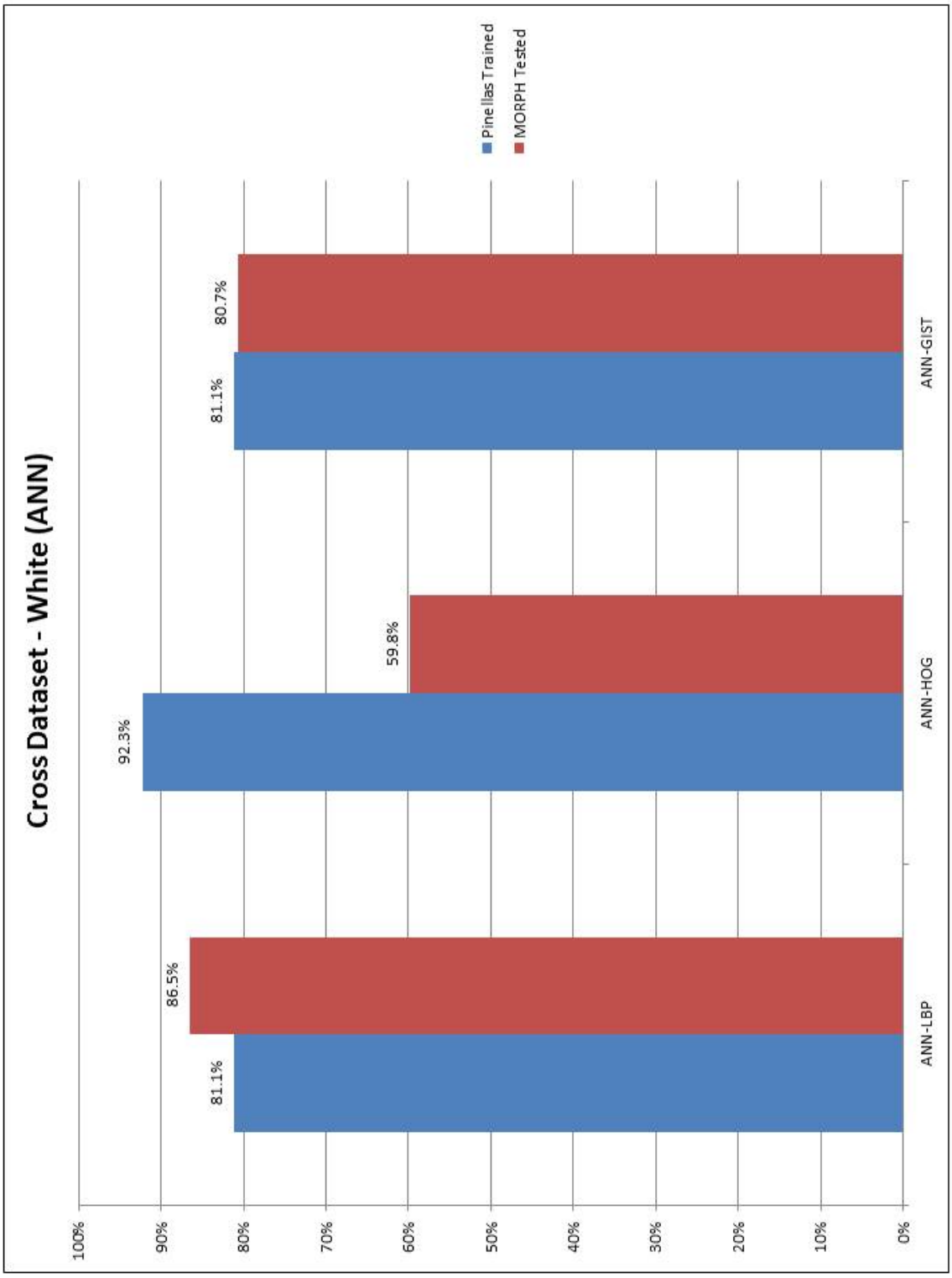

Figure 8.76: Cross Dataset - ANN White (Pinellas Trained - MORPH Tested)

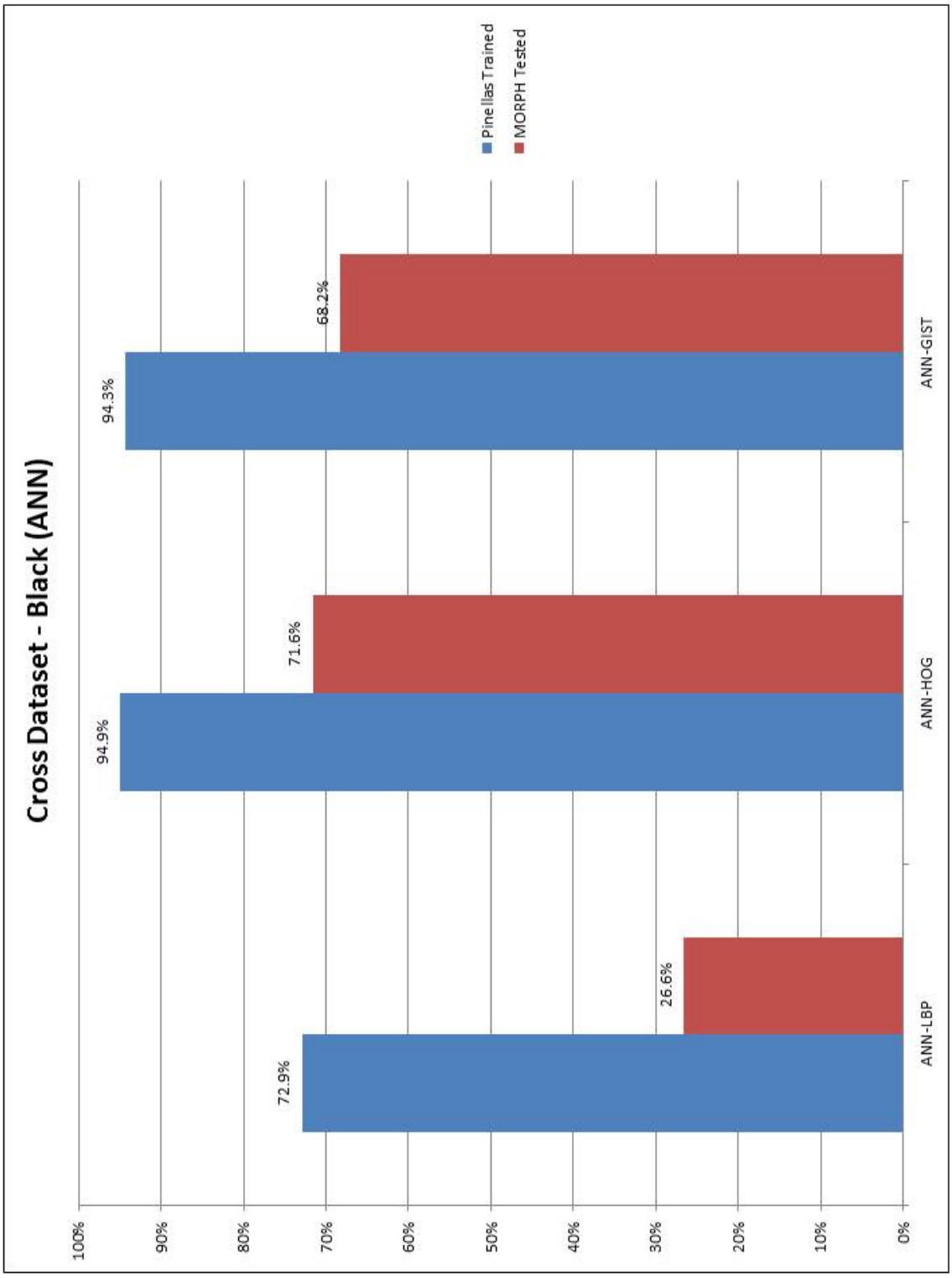

Figure 8.77: Cross Dataset - ANN Black (Pinellas Trained - MORPH Tested)

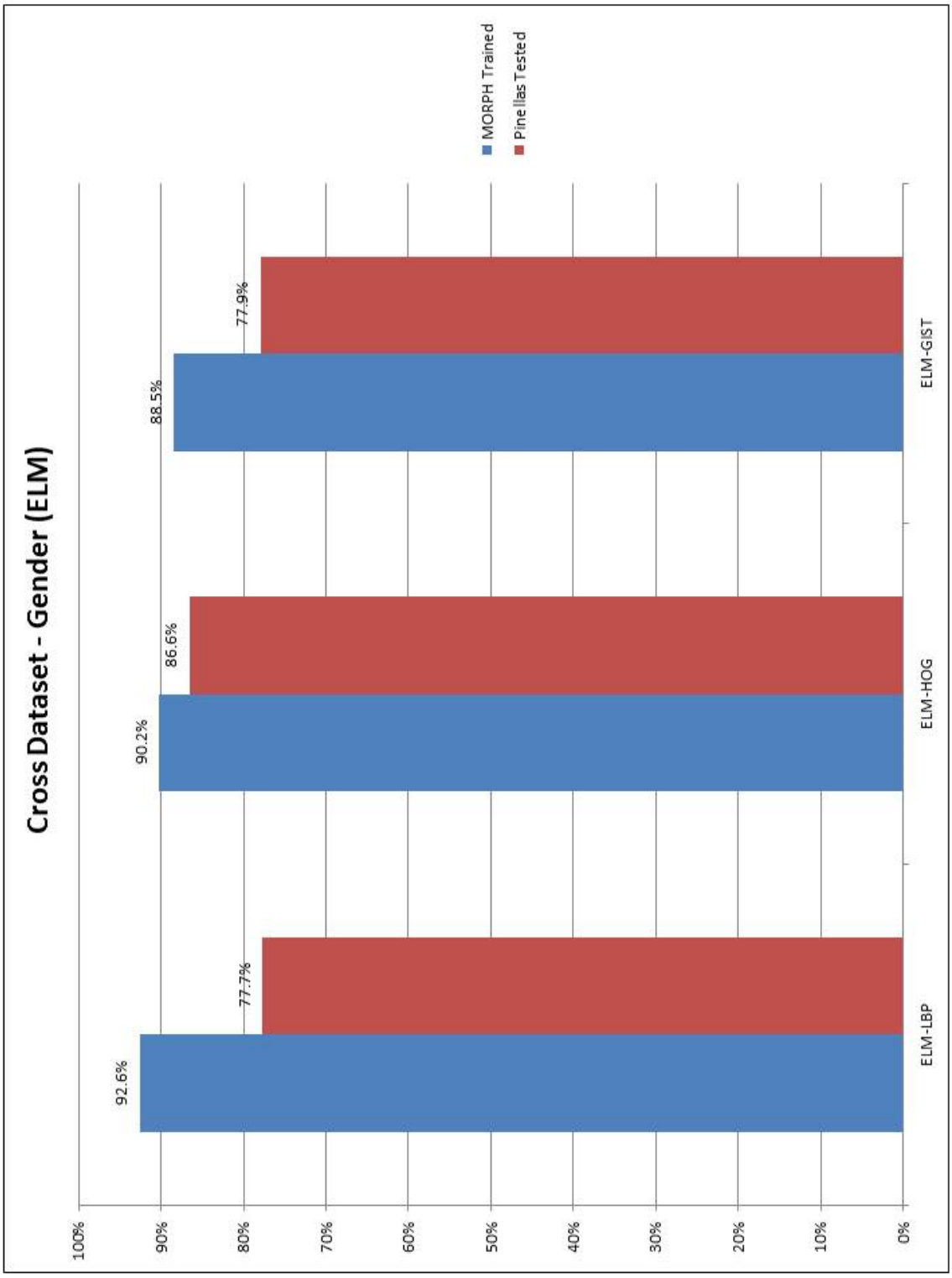

Figure 8.78: Cross Dataset - ELM Gender (MORPH Trained - Pinellas Tested)

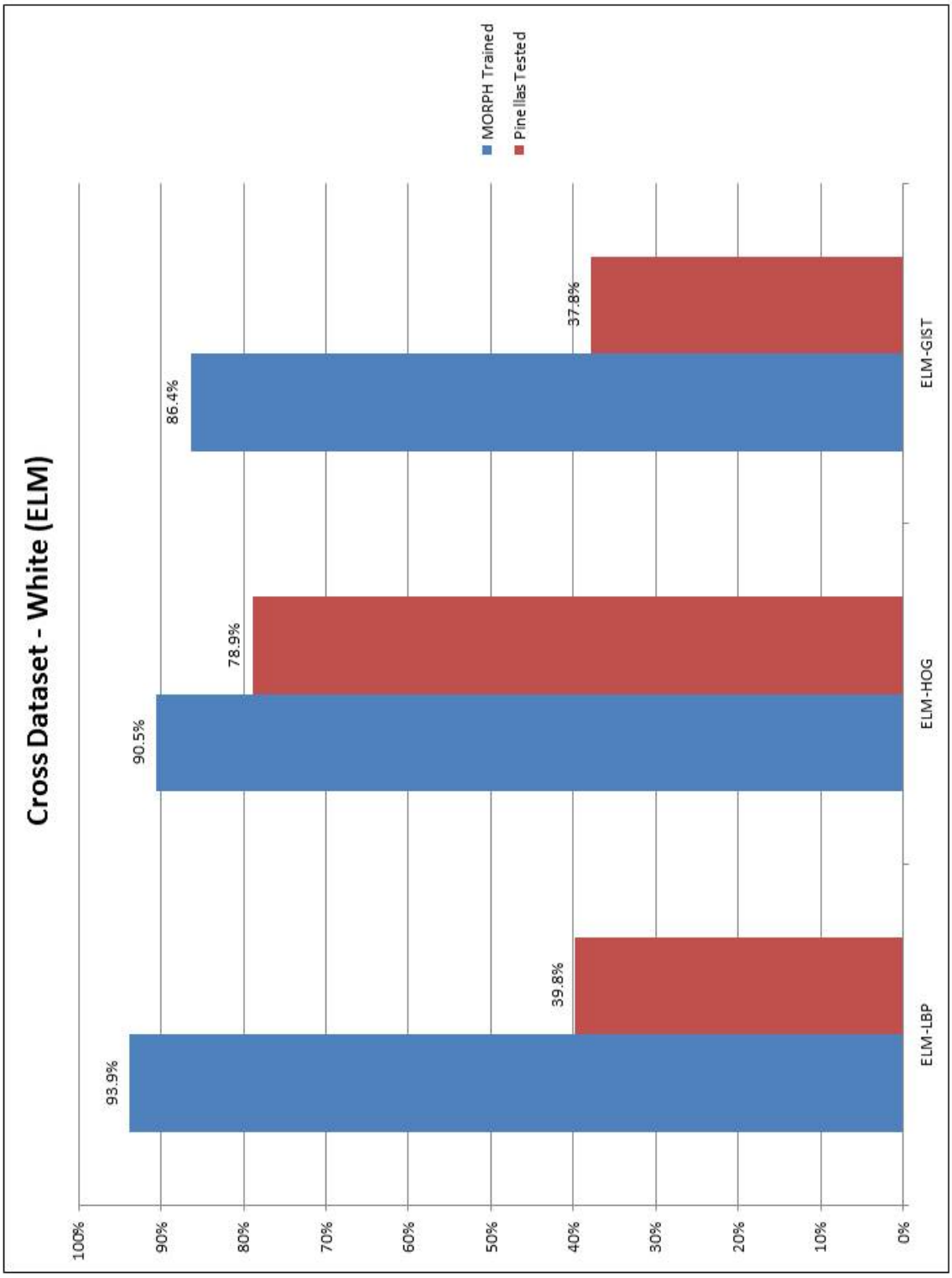

Figure 8.79: Cross Dataset - ELM White (MORPH Trained - Pinellas Tested)

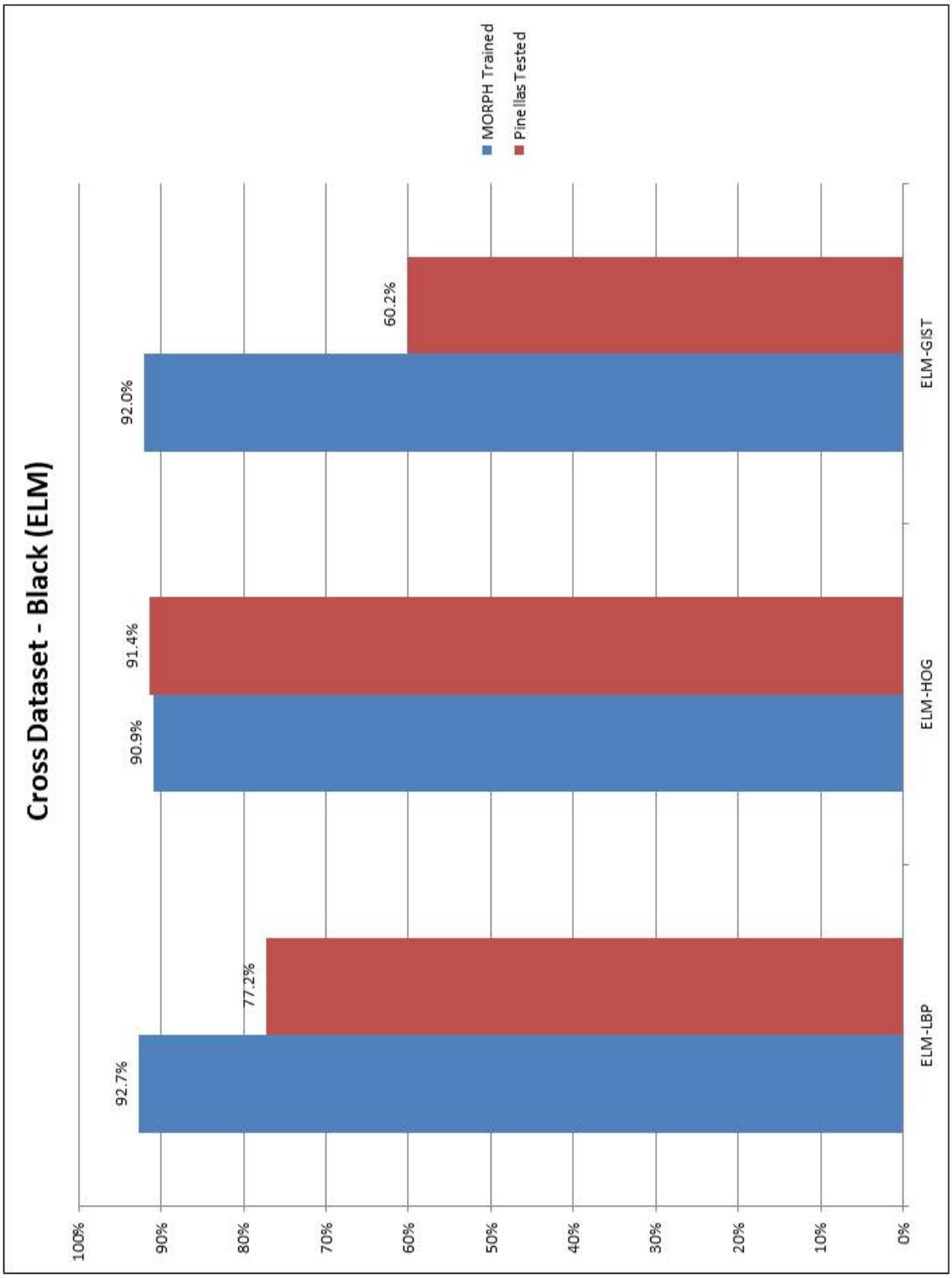

Figure 8.80: Cross Dataset - ELM Black (MORPH Trained - Pinellas Tested)

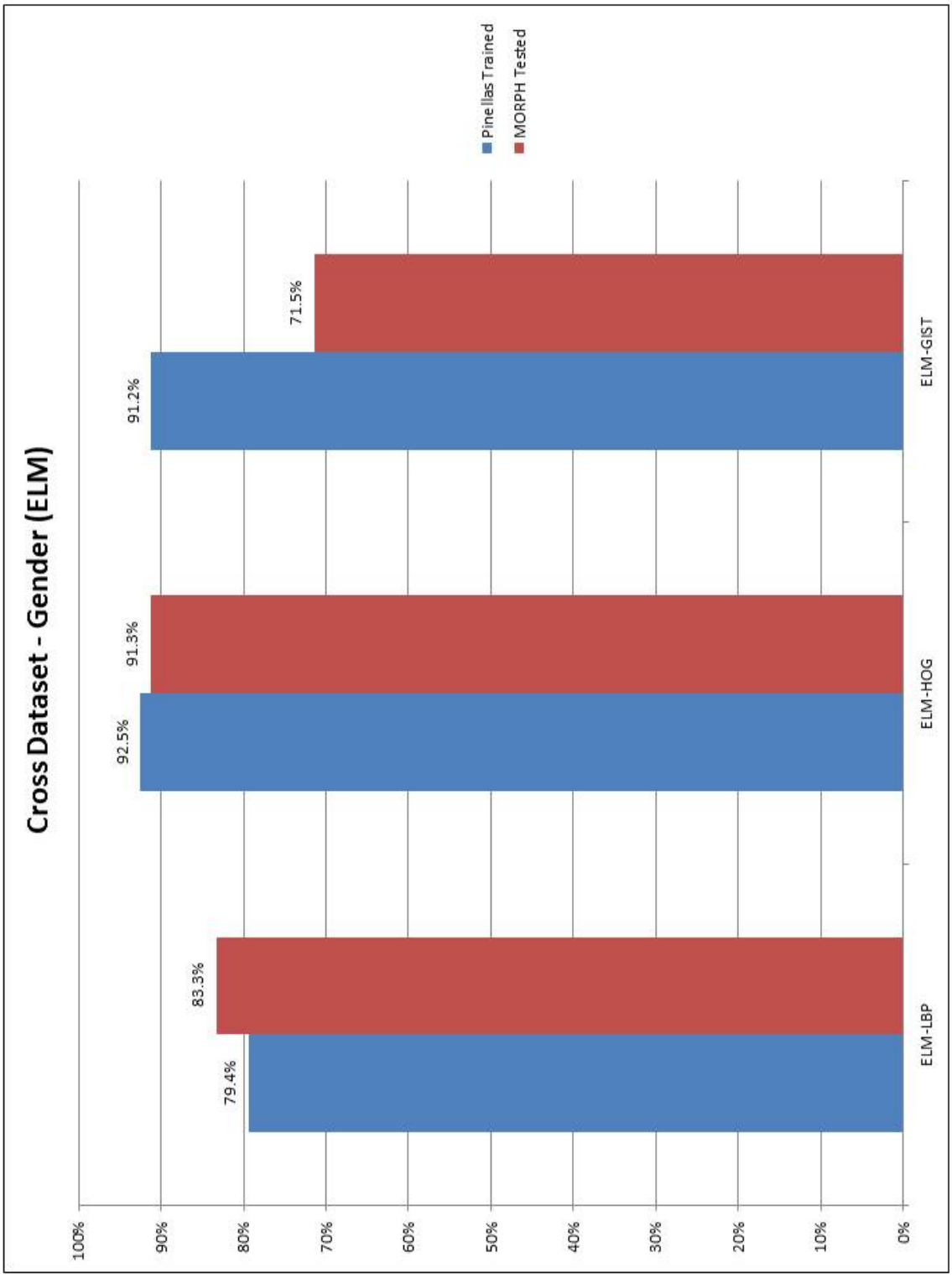

Figure 8.81: Cross Dataset - ELM Gender (Pinellas Trained - MORPH Tested)

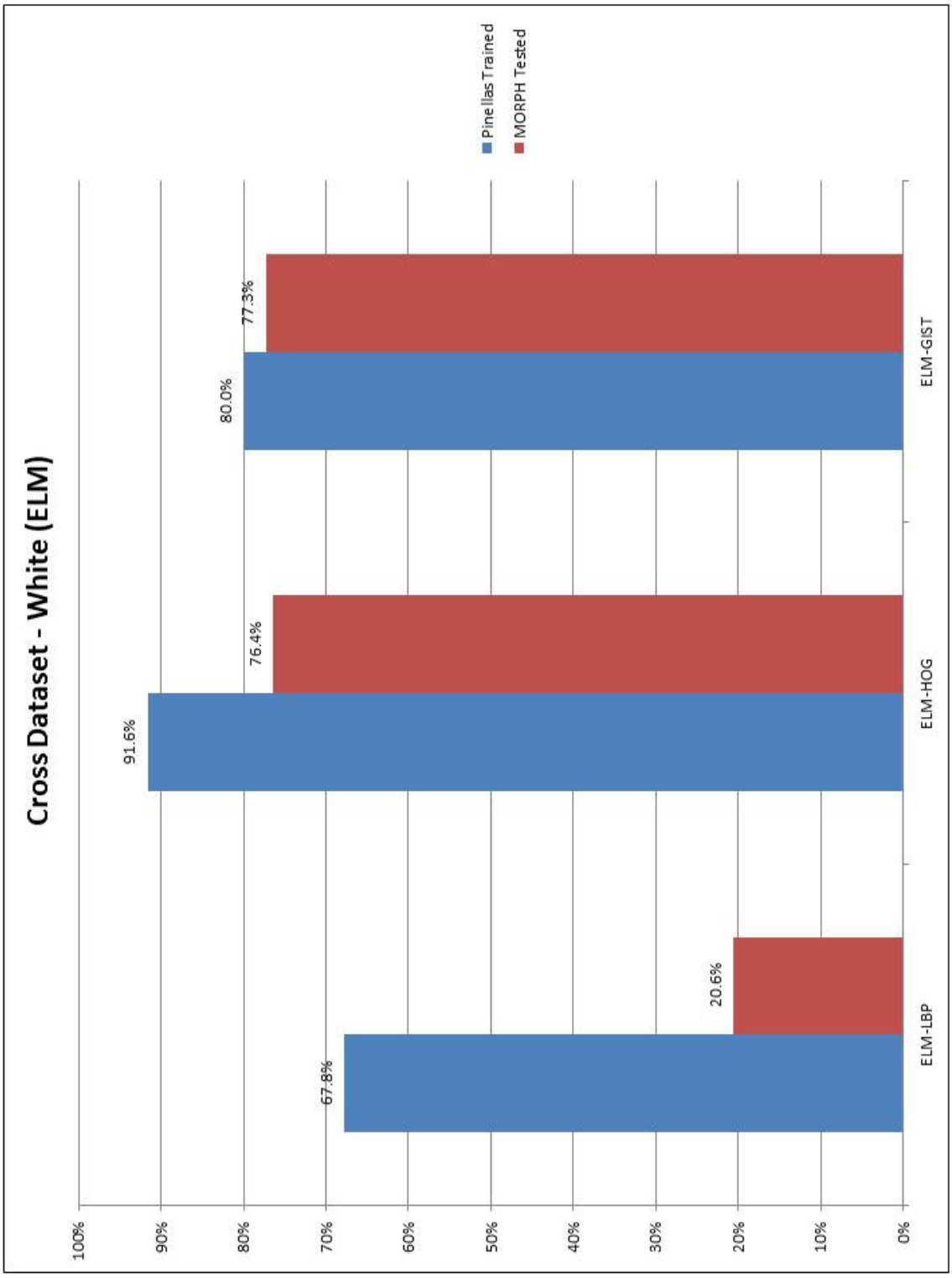

Figure 8.82: Cross Dataset - ELM White (Pinellas Trained - MORPH Tested)

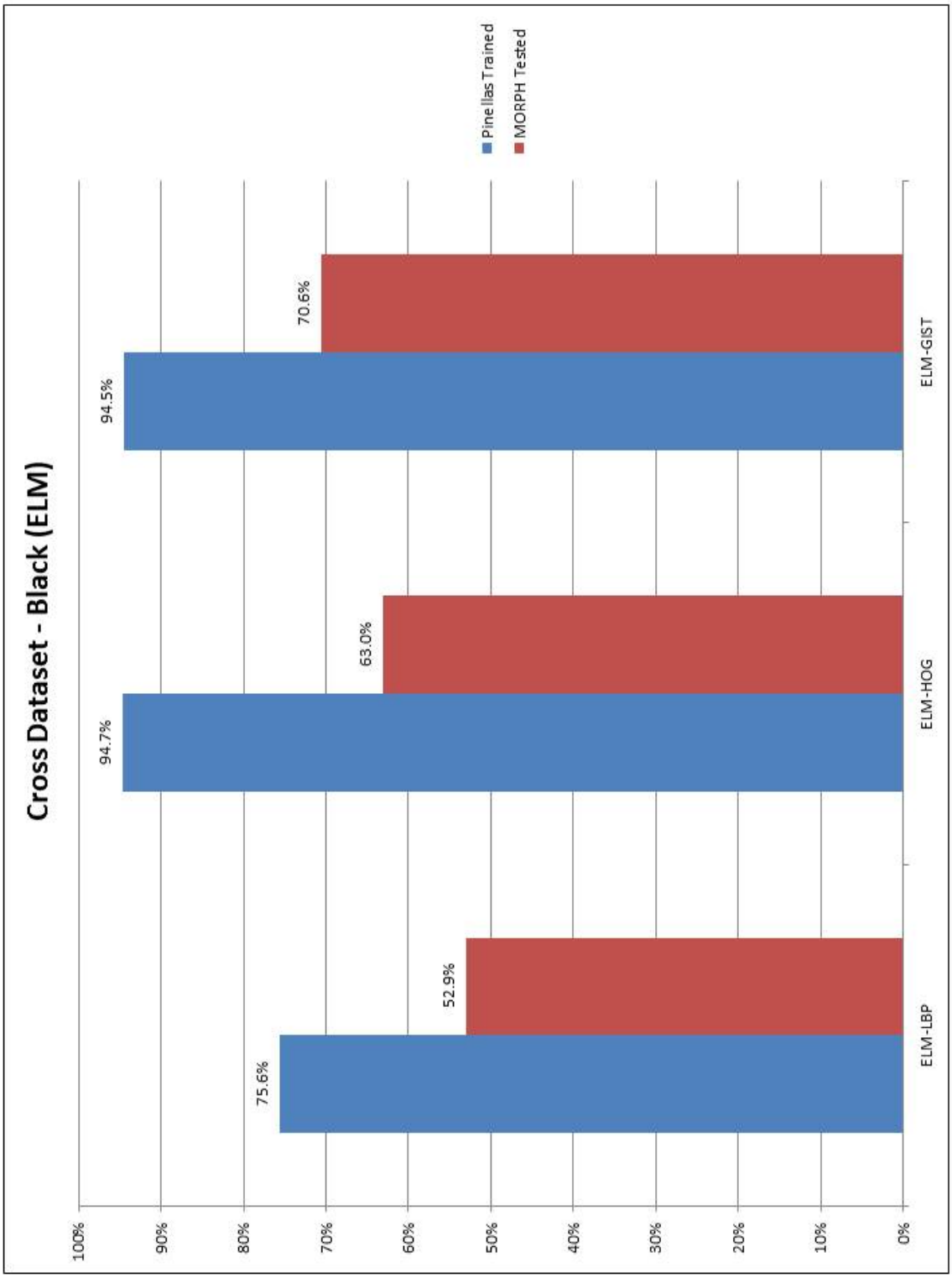

Figure 8.83: Cross Dataset - ELM Black (Pinellas Trained - MORPH Tested)