SIX NOISE TYPE MILITARY SOUND CLASSIFIER

by

Christopher Michael Shelton

BS, Mechanical Engineering,

University of Maryland Baltimore County,

2009

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Christopher Michael Shelton

It was defended on

June 25th 2013

and approved by

Jeffrey Vipperman, Ph.D., Associate Professor

Daniel Cole, Ph.D., Associate Professor

William Clark, Ph.D., Professor

Thesis Advisor: Jeffrey Vipperman, Ph.D., Associate Professor

SIX NOISE TYPE MILITARY SOUND CLASSIFIER

Christopher Michael Shelton, M.S.

University of Pittsburgh, 2013

Blast noise from military installations often has a negative impact on the quality of life of residents living in nearby communities. This negatively impacts the military's testing & training capabilities due to restrictions, curfews, or range closures enacted to address noise complaints. In order to more directly manage noise around military installations, accurate noise monitoring has become a necessity. Although most noise monitors are simple sound level meters, more recent ones are capable of discerning blasts from ambient noise with some success. Investigators at the University of Pittsburgh previously developed a more advanced noise classifier that can discern between wind, aircraft, and blast noise, while simultaneously lowering the measurement threshold. Recent work will be presented from the development of a more advanced classifier that identifies additional classes of noise such as machine gun fire, vehicles, and thunder. Additional signal metrics were explored given the increased complexity of the classifier. By broadening the types of noise the system can accurately classify and increasing the number of metrics, a new system was developed with increased blast noise accuracy, decreased number of missed events, and significantly fewer false positives.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Impulse noise generated from military testing and training events, such as demolitions, mortars, and artillery, is typically referred to as blast noise. The majority of the acoustical energy in blasts is centered between 10 and 100 Hz; therefore, these signals can travel through the atmosphere with little attenuation due to atmospheric absorption. This results in a large noise footprint, as the typical blast noise event can often be heard at distances as far as 10-20 kilometers from the source. As a result, blast noise is typically reported as the most annoying noise source around military installations [1], and has been the cause of many noise complaints [2] and damage claims. In order to better monitor and assess the noise impact on those exposed to blast noise, the military is in need of reliable noise monitors and accurate blast noise classifiers.

To date there has been some work done to develop specialized equipment for the purpose of detecting and classifying abnormal sounds in surveillance applications [[3], [4],& [5]] and other work that has specifically designed for military applications [6],[7], [8], & [9]. Cvengros et al. developed a blast noise classifier based upon sound level meter metrics using support vector machines and Bucci and Vipperman developed an Artificial Neural Network (ANN) that used statistics and other metrics related to the pressure time series. The ANN classifier was able to classify blast, wind, and aircraft from the test data with 99% accuracy using only four metrics: weighted square error, spectral slope, kurtosis, and crest factor. A prototype of the Bearing Amplitude and Measurement Analysis System (BAMAS) was then developed by Applied Physical Sciences, which included the Pitt noise classifier and could determine the approximate bearing of the event, and reject wind noise using a four microphone array [10]. Although the three class classifier (wind, aircraft, blast) had been developed, a binary (blast/non-blast) classifier was implemented.

This project is a continuation of two previous projects. The first project, directed by Brian Bucci, gathered data used to train the original UPitt classifier, examined different methods of training, and deduced which metrics would be necessary to produce the highest ANN accuracy. [6]. Researchers carefully controlled the detonation of each blast and were present for each wind and aircraft recorded to verify every sound source. Although the original UPitt classifier was 99% accurate in test data sets, a signal to noise analysis was not performed to ascertain its sensitivity to noise. The classifier's sensitivity to wind came into question when the classifier's accuracy dropped in windy conditions. In the field however, the prototype had an overall blast detection accuracy of 85%. There were a high number of false positives due to machine gun shots, thunder, and certain vehicles which weren't accounted for during algorithm training. It also continuously missed blasts with low kurtosis. This project seeks to improve upon his work by designing a new classifier able to classify more types of noise which may be confused for blast, perform a signal to noise ratio analysis, and test the classifier on samples of known wind speed to gauge its sensitivity to wind and its overall accuracy under real world conditions.

The second project which preceded this one was directed by Matthew Rhudy. He converted Brian's UPitt classifier from Matlab code into C code so that it could be implemented on a PC104 minicomputer which would operate in the field on a Bearing Amplitude and Measurement Analysis System (BAMAS) acoustic detection system. Each acoustic detection system is solar powered, continuously monitors sound, and every time a sound louder than 95 dB is heard, the system determines the time, approximate location, bearing, wind speed, decides whether or not to record the sound, and if recorded the sound is uploaded to the BAMAS website within a second after the sound is heard. Because the classifier would be working in concert with an existing system, it must be lean enough to fit onto the system's memory and fast enough as not to delay the system. This project aims to improve the accuracy of the classifier under real world conditions without significantly increasing processing time or size of the UPitt classifier algorithm.

Lastly, this project aims to increase the number of sound types capable of being classified by the ANN from three to six (blast, wind, machine gun, vehicle, aircraft, and thunder). These additional sound classes were implemented with the goal of increasing

blast accuracy by supplying more non-blast sound types for training (machine gun and thunder). Additionally, if vehicle and aircraft could be accurately classified, this new feature in conjunction with the BAMAS acoustic detection system's bearing calculation algorithms would allow for auditory tracking of vehicles and aircraft.

2.0 LITERATURE REVIEW

The six sound types being classified (blast, wind, machine gun, air craft, vehicle, and thunder) were mostly distinctive in that the qualitative sound was consistent among all sound files of similar classifications, but the time history of each individual file may vary in shape. Four methods of sound classification are: multi-layer perceptron artificial neural networks, template matching, support vector machines, and hidden markov models.

2.1 MULTI-LAYER PERCEPTRON (MLP) ARTIFICIAL NEURAL NETWORK (ANN)

An Artificial Neural Network (ANN) is a non-linear mathematical model used to find patterns between inputs and outputs. If a firm pattern can be established, an ANN can be used to predict the outcomes of certain events. Traditionally, they are composed of three parts: an input layer which contains the information given to the ANN, one or more hidden layers which use the information to predict the outcome, and an output layer which houses the predictions.

Classification of sound samples occurs in three steps. First, the sound pressure time history is recorded for each acoustic event and saved as a .csv file. Secondly, metrics are computed from the time history, linear spectrum, and log spectrum of every acoustic event. Lastly, those metrics are used as inputs by classifier to determine the acoustic event source. Figure 1 displays how the the input vector x enters the neuron of the hidden layer, is multiplied by the weight matrix W, summed with the bias vector b, squashed with an activation function a to ensure that the hidden layer's outputs are between a desired range of values, and sent to the output layer which yields the output vector y.

Figure 2 shows how the output of each neuron comes together to form the predictions in a multi-hidden-layer ANN, also known as a multilayer perceptron (MLP). The first column of boxes represent three inputs being entered into the MLP while the lines coming from them represent the path each input follows. Each input is fed into every node of the first input layer where it is multiplied by a weight, added with a bias, and summed with every other input before heading to the squashing function. Lines coming from each node in the first hidden layer represent the outputs of the first layer being fed into every node of the second hidden layer. There, the numbers undergo the same process as the inputs experienced in the first hidden layer where they are multiplied by weights, added with biases, summed, and sent through a squashing function. This process repeats until numbers emerge from the output layer's squashing function; numbers which represent the ANN's prediction for the sound source.

An ANN can have as many hidden layers with as many neurons as needed. Complex ANN structures greatly increase the ANNs ability to handle complicated tasks; however, such ANNs are computationally expensive, require longer training times, and may need more training samples. Once an ANN is trained, "Network Pruning" (the systematic removal or addition of hidden layers and neurons) is applied to ensure that only the essential number of hidden layers and neurons are used.

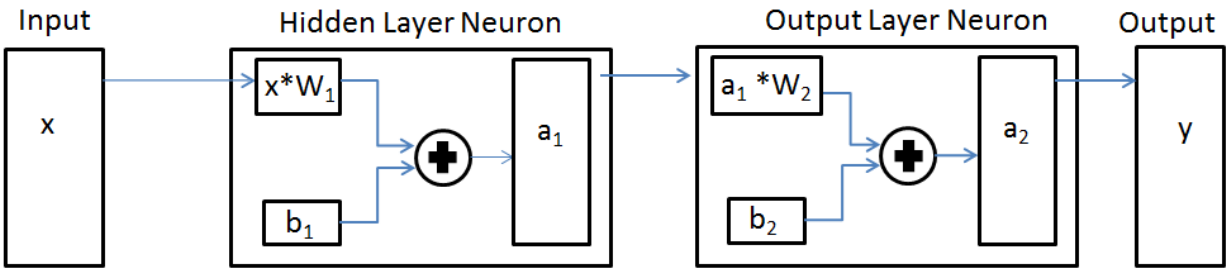

Figure 1: Neuron Structure

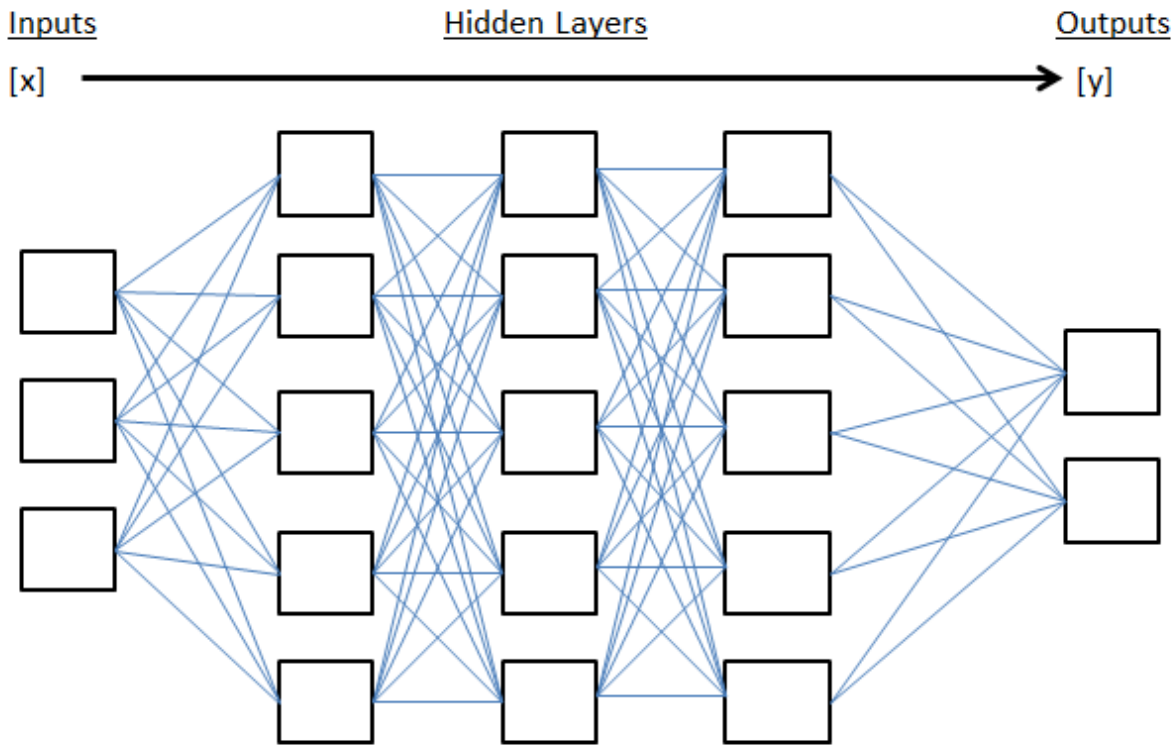

Figure 2: Multi-Layer ANN Structure

Bucci (2007) developed a MLP based classifier to classify sound into three categories: blast, aircraft, and wind. The author also developed a Bayesian classifier which binarily categorized sound files as either blasts or non blasts [11]. It performed on par with the MLP classifier but ultimately the MLP was easier to implement.

Sound files used during the MLP ANN classifier's development came from the SERDP library which will be discussed later in this work. Of those files, two-thirds were used for training, one-sixth for validation, and the remainder were used for testing. Four metrics were utilized: weighted square error, kurtosis, spectral slope, and crest factor. Bucci also compared the MLP based classifier with another ANN classifier based on support vector machines with radial basis function kernel. The MPL based classifier achieved 100% classification accuracy while the support vector machine based classifier achieved a classification accuracy of 95%[6].

2.2 TEMPLATE MATCHING METHOD

Template matching is a non-neural method which compares input vectors to template vectors of known origin, computes the correlation coefficient of the two vectors, and assigns classifications depending on a set threshold of acceptance for that correlation coefficient. Rodriguez (2008) developed a gunshot detection algorithm based off of the template matching method. The algorithm was to be utilized to prevent illegal hunting in tropical rainforests. Although other classification algorithms were available, the author was restricted to low complexity classification algorithms due to hardware restrictions [3] .

Although many factors affect the sound propagation of gun shots, the largest factors were acoustic surroundings such as temperature, wind speeds, foliage density, air moisture, distance from the sound source, and soil characteristics [3]. To account for these factors, sound samples were collected from a variety of locations within a dense tropical forest at a 48 kHz sampling rate with a high quality digital recorder. Firearms of five different calibers were fired at distances of 30, 90, and 250 meters from the sound recording equipment. Additional examples of non-gun shot samples such as chain saws, planes, birds singing, rain showers, streams, wind, human voices, and mat-lab generated white noise were also recorded/generated. Sound pressure levels ranged between 90 dB and 98 dB. The signal was then compared to two gunshot template vectors (one recorded at 30 meters and another at 90 meters from the noise source[3]). The classifier was then evaluated using 45 positives (samples with gun shots) and 15 negatives. Both template vectors produced a 91% true positive accuracy and 0 false positives.

2.3 SUPPORT VECTOR MACHINES (SVM)

Support Vector Machines are algorithms which analyze data, detect a pattern, and sort data according to two classes (a positive example and a negative example). In theory, there exists a hyperplane between the best data vectors and their associated positive or negative examples. A hyperplane can be approximated as $w \cdot x + b = 0$ where x is the data vector, w is the weight of the hyperplane, and b is a bias[5]. The separation of positive and negative examples can be described by $w \times x_i + b > 1$ if $y_i = 1$ and $w \times x_i + b < -1$ if $y_i = -1$.

During training, weights of the hyperplane are computed so that the distance between the negative and positive example outputs is maximized. Points within the data vector which contribute the most to the hyperplane separation (i.e. the most important points) are called *support vectors*. There may be positive cases which yield negative results. For such exceptions, a cost function can be introduced which will augment the hyperplane optimization to account for such irregularities. If the data can't be linearly separated into positive and negative examples, *kernel functions* are used to transform the coordinates of the non-linear data set into a linear k-dimensional space. Figure 3 is displays how a polynomial kernel function performs the transformation[5].

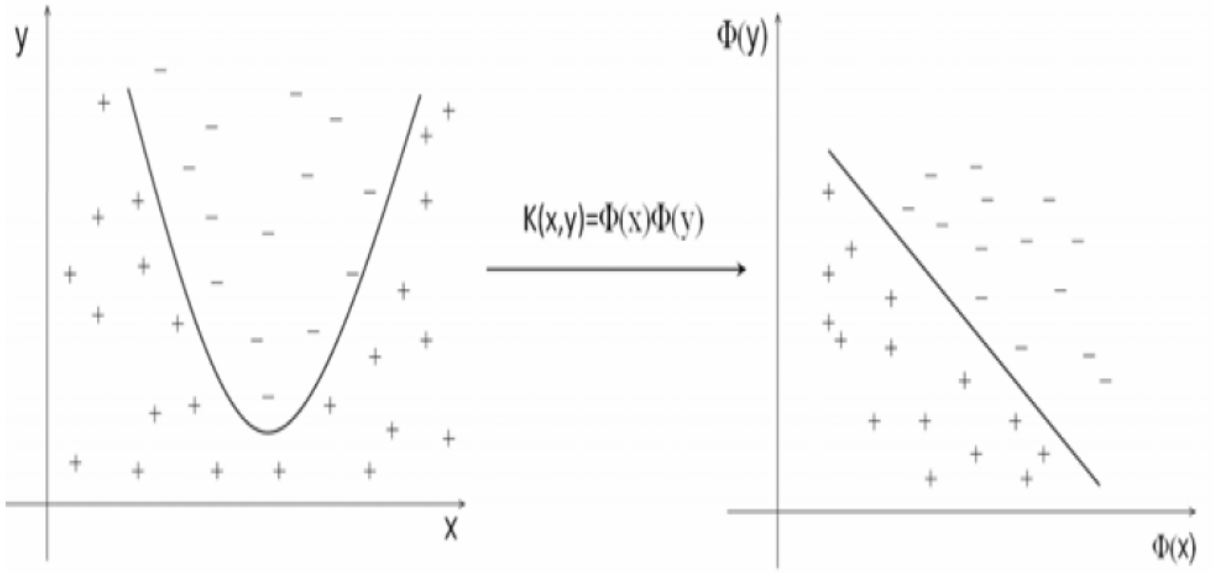

Figure 3: Polynomial Kernel Hyperplane Transformation

Lopatka (2010) developed a support vector machine based sound classifier to sort sounds into dangerous and non-dangerous categories to supplement visual surveillance systems [5]. A total of 28 metrics were used as input data which consisted of nine (9) energy parameters, two (2) transient-sensitive parameters (sound length and ratio of sound loudness to background loudness in decibels), and seventeen (17) MPEG-7 features including but not limited to audio spectrum envelope, spectral flatness for different bands, spectral flatness for mean variance, audio spectrum spread, and others.

Dangerous sound files available were explosions (16), broken glass (120), gunshots (157), screams (26), and other non-dangerous sounds (51) totaling to 372 sound files. Eighty percent (80%) of the data was used for training while the remaining was used as a test data set. When trained, the support vector machine based classifier correctly classified 95.38% of the test sound files. When Gaussian white noise was added to the test files, system accuracy saw a high of 90.77% at an SNR (dB) of 30, and a low of 73.85% at an SNR (dB) of 10 which implies a significant degree of noise sensitivity.

2.4 HIDDEN MARKOV MODEL (HMM)

Hidden Markov Models are produced by constructing patterns from observed and hidden states within a system, and saving those patterns as probability distributions. They are considered the simplest dynamic Bayesian network and are commonly used for temporal pattern recognition for speech, hand writing, and audio surveillance [[12],[13]].

Chan (2010) developed a HMM classifier to sort sound into three categories: human screaming, non-human emergency sounds such as breaking glass, explosions, or gun shots, and background noise for all other noise samples. Seven acoustic metrics were calculated for each sound file: weighted average delta energy, LPC spectrum flatness, FFT spectrum flatness, zero crossing rate, harmonicity, mid-level crossing rate, and peak/valley count rate. The HMM used an 8-state parallel left-right model with three (3) incoming and outgoing connections in each state [4]. Three (3) separate HMM's were trained: one for screaming, another for non-human emergency sounds, and one last HMM for background noises. 400 human voice files (which contained both screams and regular speech), 300 non-human emergency sound files, and 100 background/ambient sound files were used in training.

Probabilities of each sound class were produced by feeding input feature vectors containing the acoustic metrics computed from unknown sound files into each HMM. A decision rules and duration labeling module was designed to decide the source of the sound class using the three probabilities and a set of sensitivity thresholds. Each threshold was derived during training under various SNR conditions using Gaussian white noise. Ultimately, three threshold sets were derived: high (SNR $<$ 10 dB) sensitivity, medium (10 dB $<$ SNR $<$ 20 dB) sensitivity, and low (SNR $>$ 20 dB) sensitivity. Users were allowed to select the sensitivity setting they require for their needs.

The human/non-human sound classification accuracies of the Chan classifier are displayed in Table 1. It was 94.5% accurate at an SNR of 30 dB; however, the accuracy fell to 81.3% at an SNR of 5 dB. Table 2 shows that at an SNR of 30 dB the classifier had a 18.4% human scream false positive rate and a 16.3% emergency sound false positive rate which only gets worse as the SNR decreases. Although the false negative rates shown in Table 3 were relatively low, the high percentage of false positives do not inspire confidence in the HMM method.

Table 1: Classification Error of the Human/Non-Human Sound Classifier.

SNR (dB)	Error Rate (%)
30	5.5
20	8.3
10	13.5
5	18.7

.

Table 2: Chan Classifier False Alarm Rate.

SNR (dB)	Human Scream Detection False Alarm Rate (%)	Emergency Sound Detection False Alarm Rate (%)
30	18.4	16.3
20	19.5	19.2
10	22.1	20.5
5	23.6	25.8

.

.

Table 3: Chan Classifier Miss Rate.

SNR (dB)	Human Scream Detection Miss Rate (%)	Emergency Sound Detection Miss Rate (%)
30	6.6	7.6
20	8.3	8.2
10	10.2	10.1
5	13.9	14.2

2.5 SUMMARY OF CLASSIFICATION METHODS

There was no compelling reason to use a different classification method other than the MLP ANN, as will be seen in the following comparison. Although Rodriguez's classifier's true positive accuracy was high and its false positive rate low, the negatives used during the classifier's evaluation were few, sounded nothing like gun shots and were inclusive only of the typical sounds likely to be heard in tropical rain forest environments. Chain saws and chirping birds may be common in the locations where his classifier was planned to be used and the Rodriguez classifier's ability to discern them from gun shots proves its use at those location; however, the exclusion of more impulsive negatives does not inspire confidence in the classifier's accuracy in urban environments.

Secondly, Rodriguez was limited to a low power VLSI circuit which ruled out more powerful classification algorithms. The existing hardware utilized in the study directly preceding this was was a PC104 single board computer with an Intel XScale PXA255 CPU which can handle more complex algorithms [10]. Thirdly, blast time histories have a fairly distinctive shape, but blasts mixed with wind have amorphous time histories like wind. Such shapes are highly unideal for template matching and would not decrease the high number of wind caused false negatives reported at sites implementing the currently installed UPITT classifier.

The Lopatka support vector machine classifier had high accuracy (95%). Unfortunately when noise was added to the signal, the accuracy dropped dramatically; indicating a relatively high noise sensitivity. Considering that noise is always present in

real environments and the system sees a five percent (5%) reduction in accuracy even at high SNR ratios, extensive noise reduction methods would be necessary if this system were to be implemented. The UPITT classifier can sometimes be situated in windy locations; requiring a system which is particularly insensitive to noise.

Chan's HMM classifier achieved 94.5% accuracy at a SNR of 30 dB (4.3% higher than the Lopatka classifier). Although different sets of data were used in training and evaluation, both had the same goal: classification of dangerous sounds. Bucci compared an ANN based on support vector machines and another based on MLP under similar circumstances and found that the later was the superior classification method. In both cases, support vector machine based classifiers were designed for similar means and were proven to be the less effective method.

Bucci's MPL classifier was 100% accurate (4.5% higher than the Chan classifier). The MLP classifier was trained using roughly two times the number of files as the HMM. It appears that of the four classification methods examined, an MLP ANN is still the best solution; however, an SNR analysis must be performed to test it's sensitivity to noise.

3.0 SOUND CLASSES

The proposed sound classifier has six noise targets which will be described, along with salient features used during human classification.

3.1 IMPULSES

Impulses (informally known as blasts) occur when shells or explosives are fired or detonated. Figure 4 displays the sound pressure time history, spectrum, and log spectrum from top to bottom, respectively, of a typical blast sample. The sound pressure remains steady until the blast occurs; causing a steep rise in pressure followed by a deep trough which quickly stabilizes at the ambient level. Most blasts have a kurtosis value greater that 10 from top to bottom, respectively. Most of the sound energy is centered between 10-50 Hz, depending on the size of the ordinance.

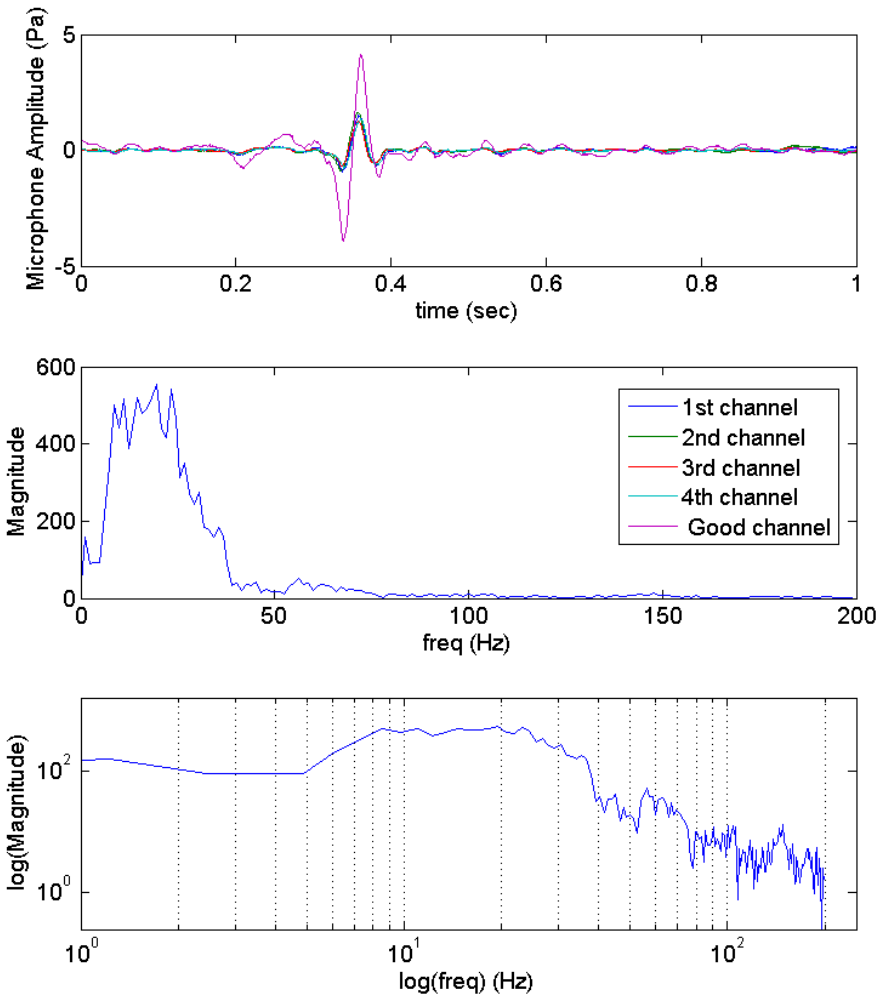

Figure 4: Typical Blast Sample Graphs

3.2 WIND

Wind sample time histories typically are amorphous with kurtosis values averaging around 3. Theoretically their log spectrum usually contain a straight sloping line across all frequencies; however, the straight line is only seen between 10 Hz and 200 Hz, due to the low-frequency roll off of the measurement system. The linear spectrum graph in Figure 5 shows that wind has a large amount of energy at very low frequencies (< 20 Hz).

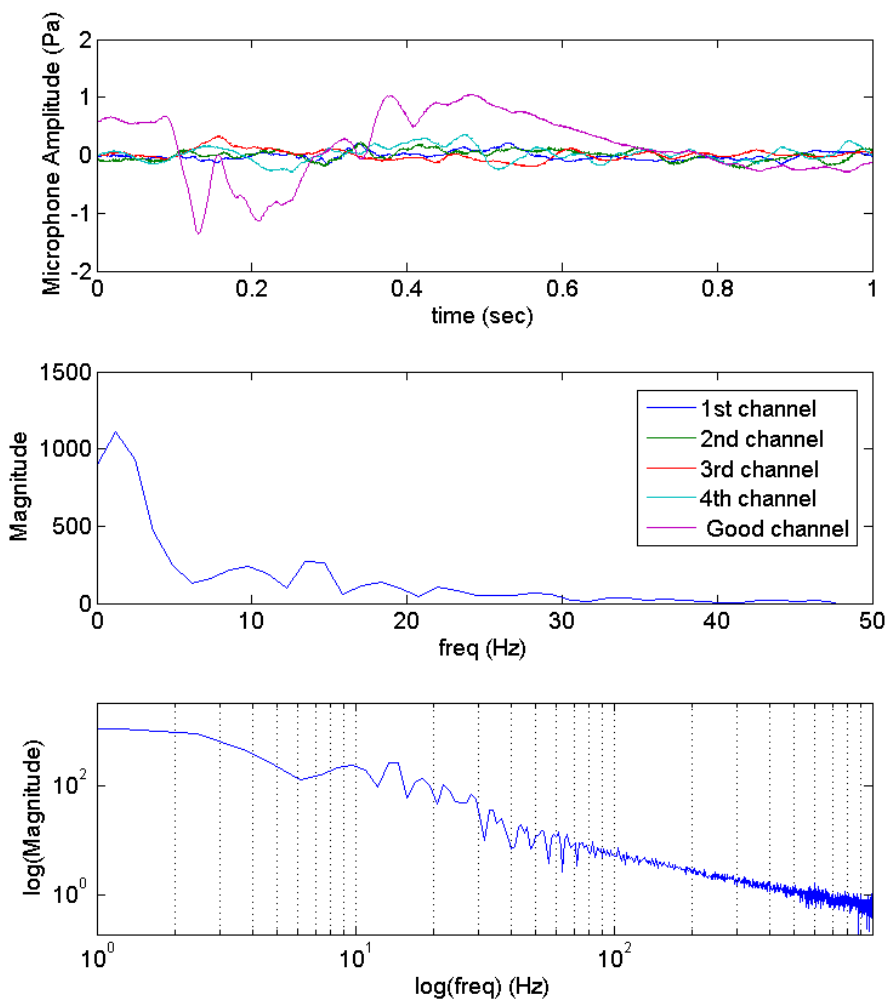

Figure 5: Typical Wind Sample Graphs

3.3 MIXED BLASTS

Some blasts occur while a strong gust is blowing. The effect is a "mixed" blast, which shares the sudden sound pressure increase of a blast, but the amorphous properties and low kurtosis of wind. Figure 6's linear spectrum graph shows a large energy spike between 0 and 20 Hz like a wind sample and a smaller yet prominent energy spike between 30 and 50 Hz like a blast sample.

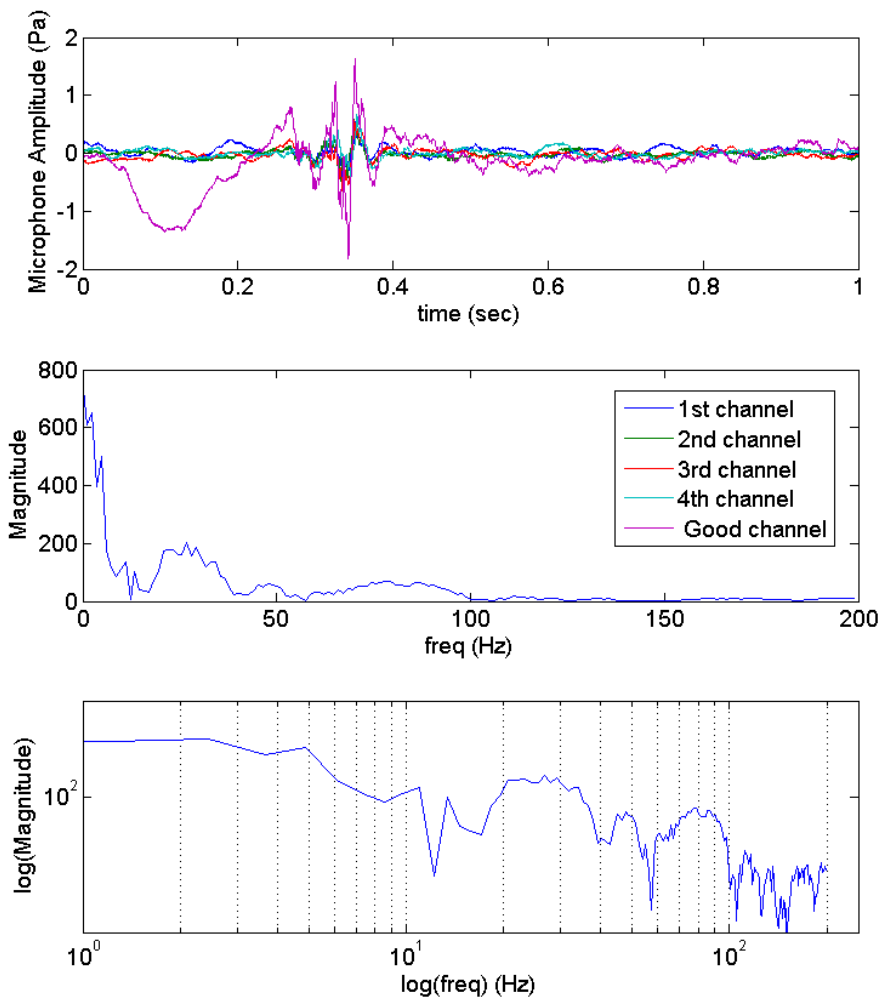

Figure 6: Typical Mixed Blast Sample Graphs

3.4 MACHINE GUN

Like blasts, machine gun shots seen in Figure 7 cause a sharp increase in sound pressure followed by a sharp but much more rapid decrease, then stabilization. Also like blasts, machine gun shots may have a high kurtosis. The differences are that the sharp increases and decreases of machine gun shots occur less gradually than those of blasts. Machine gun shots typically occur in bursts of two or three while blasts occur as single events. Their linear spectrum graphs show broad energy spikes at much higher frequencies than blasts or wind while the log spectrum graphs show larger amounts of high frequency noise than blasts of wind.

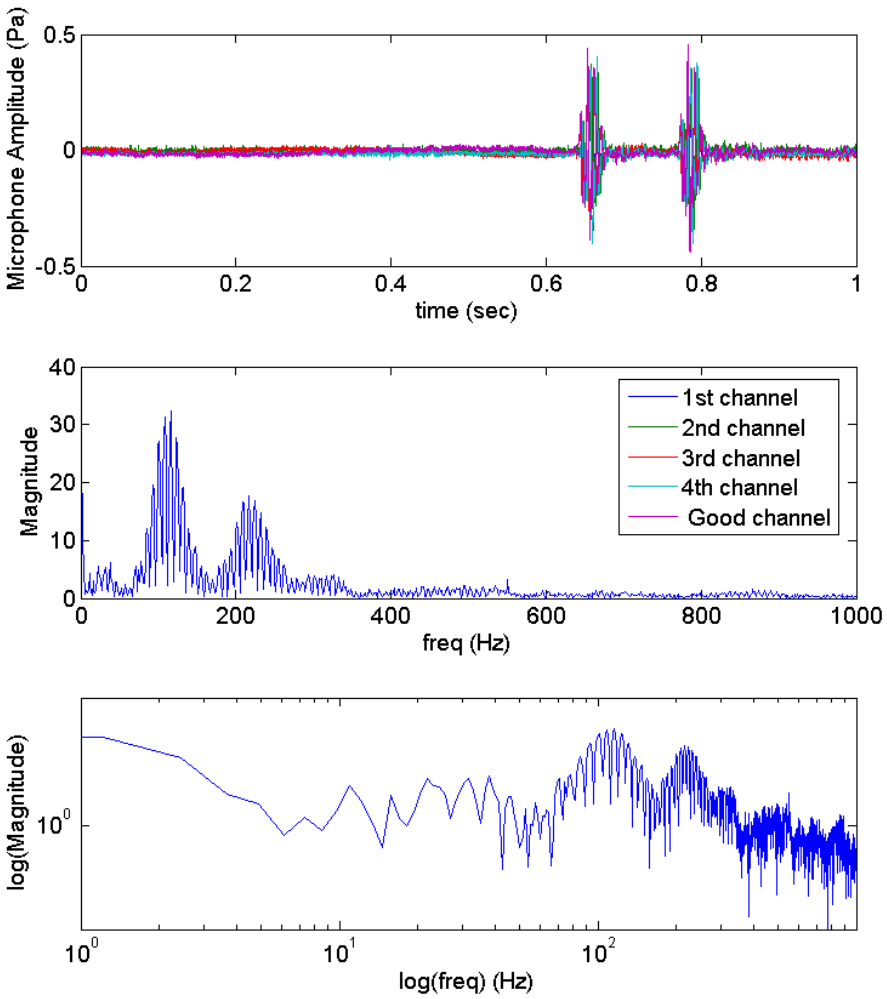

Figure 7: Typical Machine Gun Sample Graphs

3.5 VEHICLE

Vehicle sound pressure time histories in Figure 8 usually have a periodic shape with corresponding vertical lines in the spectrum graph. Those lines occur at the operational and harmonic frequencies of the engine. Occasionally motorcycles will backfire; causing blast false positives.

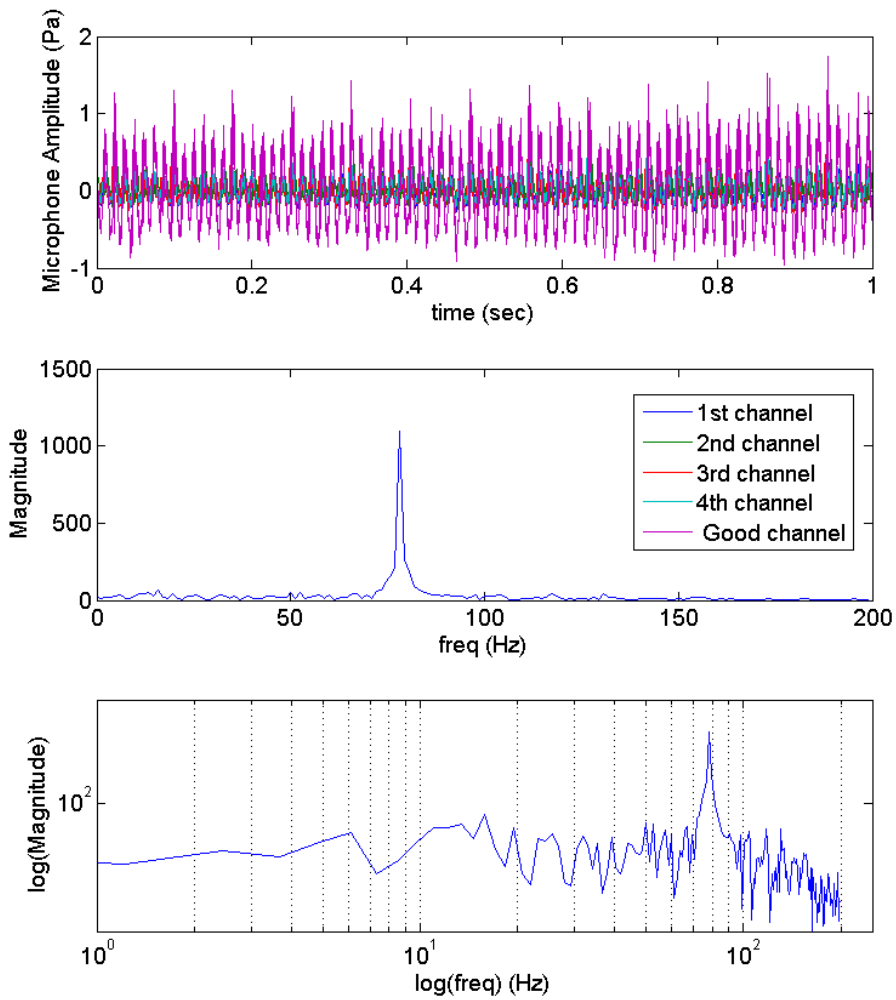

Figure 8: Typical Vehicle Sample Graphs

3.6 AIR CRAFT

Some aircraft, such as helicopters in Figure 9 and small single engine planes, are very similar to vehicles in that they have periodic shapes and vertical lines in the spectrum. Jets have periodic shapes, but at much higher frequencies as seen in Figure 10. They also have turbulent noise energy centered between 400 and 1000 Hz, which makes visually confirming the periodic shape significantly more difficult. Turbulent noise energy also manifests in the spectra as dense distributions of energy between 400 and 1000 Hz. Although all three machines are technically forms of aircraft, the differences between jet from small single engine planes and helicopter are large enough to consider them as two separate classes for ease of classification. Considering how similar small plane and helicopter time histories and spectra are to those of vehicles, the three sound sources could be lumped together and called simply "vehicle."'. When propeller plane and helicopter and lumped with vehicle like this, classification accuracy was found to improve by a few percent.

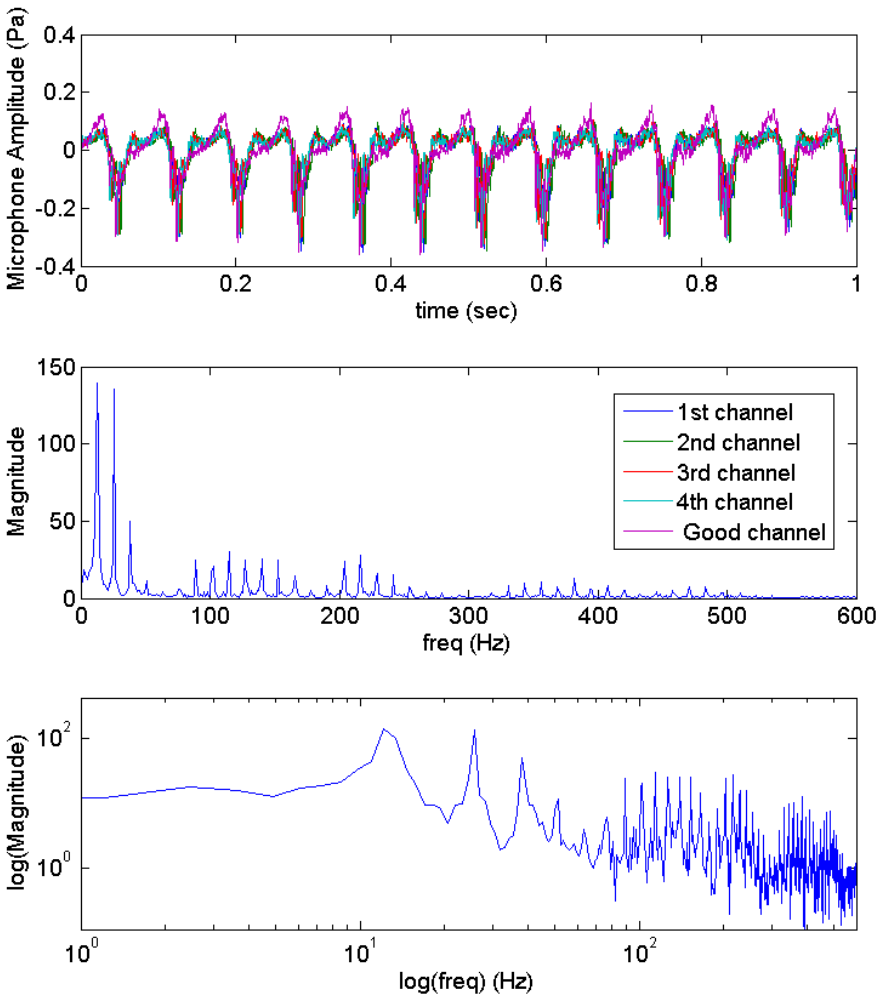

Figure 9: Typical Helicopter Sample Graphs

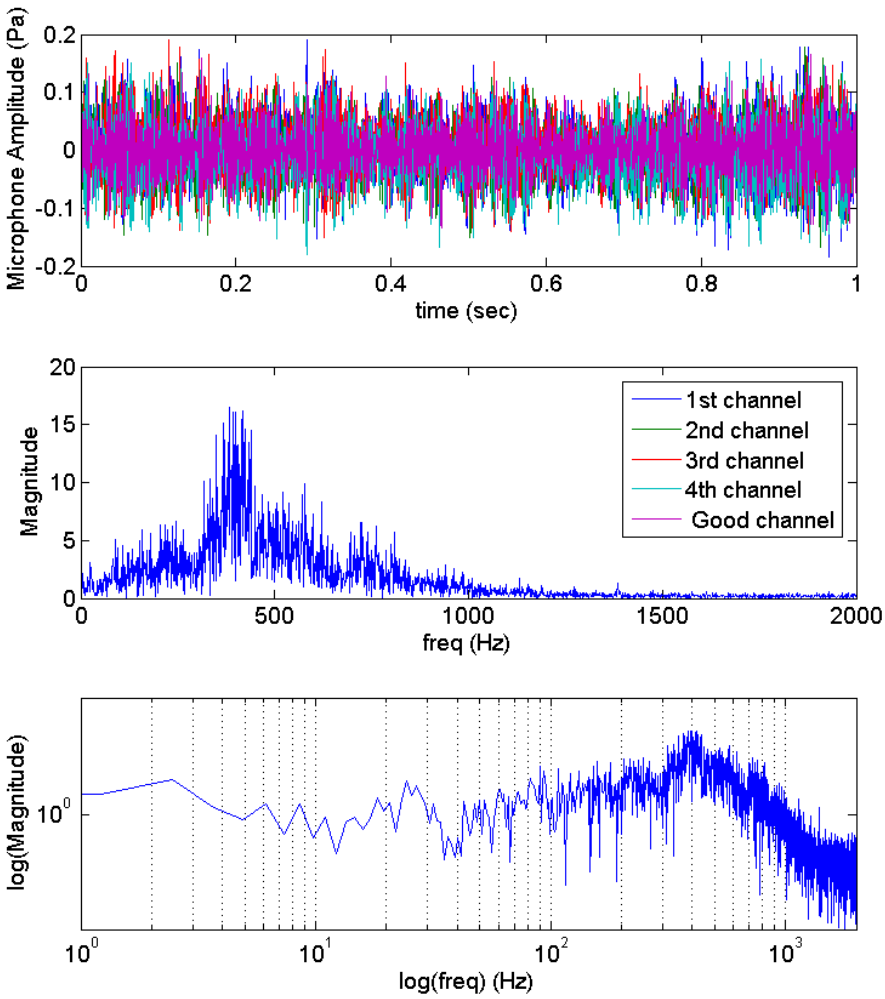

Figure 10: Typical Jet Sample Graphs

3.7 THUNDER

Thunder is an exceptionally tricky class to properly identify. It sounds like a blast, occasionally has high kurtosis, but it's time history is less uniform and the spectrum (shown in the bottom two plots of Figure 11) has more energy in higher frequencies than a typical blast.

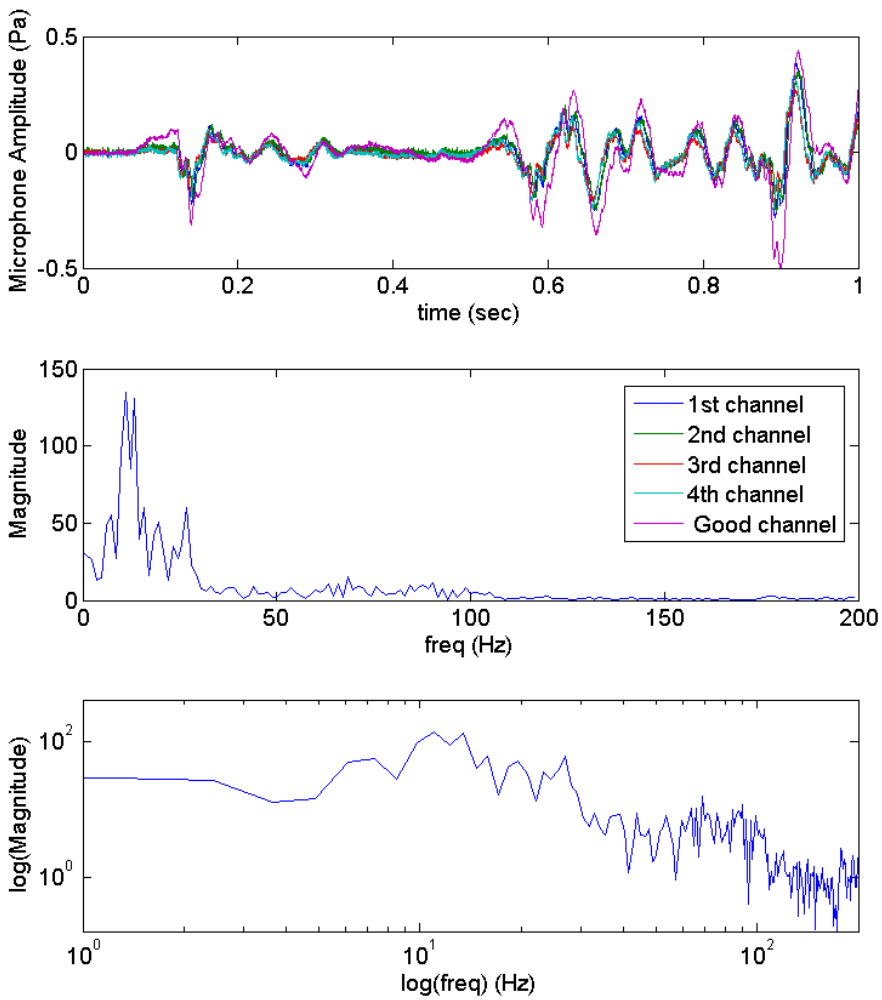

Figure 11: Typical Thunder Sample Graphs

4.0 DATA COLLECTION

Two libraries of recordings were used for the study. The existing SERDP Library was developed during the initial data collection stages of the original noise classifier project while the BAMAS Library was collected using BAMAS prototype microphone arrays[9]. During Bucci and Rhudy's works, the original UPitt classifier was trained using SERDP data, installed on the BAMAS arrays as a binary blast classifiers, and deployed at three military bases [[14], [15], and [16]].

4.1 SERDP LIBRARY

A Bruel and Kjaer 4193 infrasonic microphone with a bandwidth of 70 mHz to 20 kHz, as represented in Figure 12, was used to record the sound pressure data with a sampling frequency of 10 kHz [6]. Specific wind speeds and direction ranges weren't recorded but some gusts reached speeds as high as 50 mph. A six inch wind screen was used to block lower velocity wind noise.

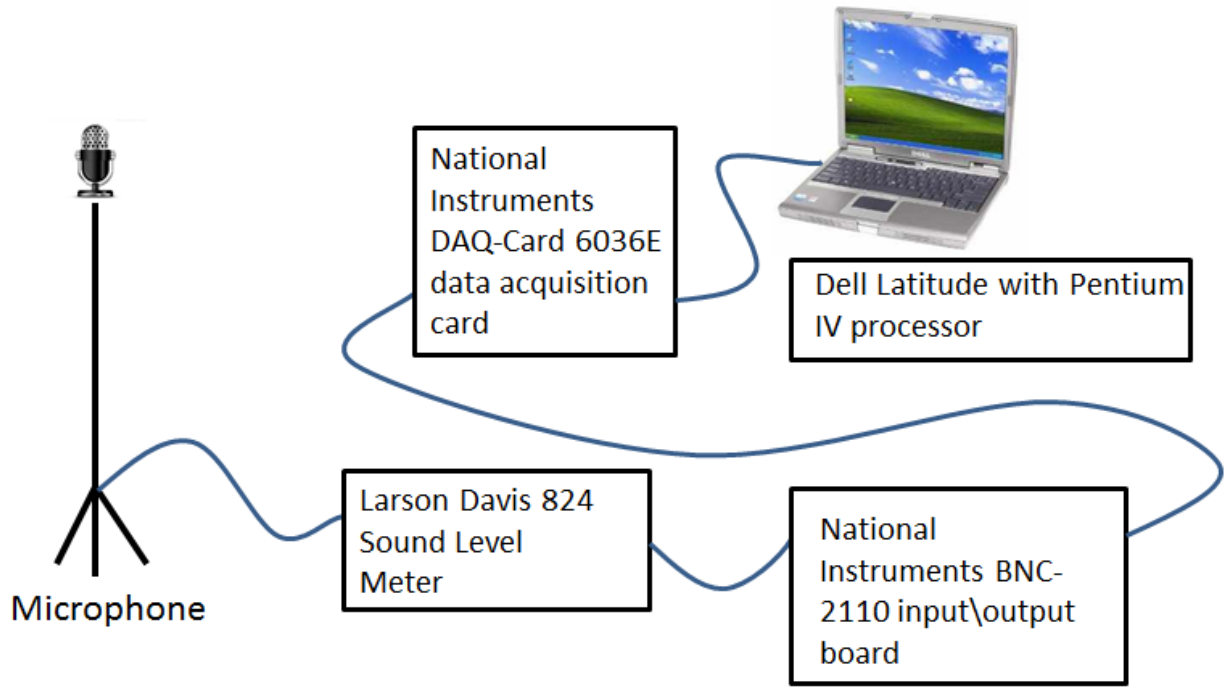

Figure 12: SERDP measurement setup

The SERDP library of 954 waveforms contains records of certain ideal acoustic events: blasts, wind, and aircraft. Of the 954 waveforms, there were 278 blasts, and 676 non-blasts. Blasts were recorded between 0.5 and 8 km from their noise sources. Measurements were taken at a variety of different locations to account for the terrain's effects on waveform measurements. The terrains examined were flat open plains, mountainous dense forests, and other various locations. Using the data from the library, the original classifier was developed using only four metrics: kurtosis, weighted square error, spectral slope, and crest factor. It obtained 99% accuracy for blast, wind, and aircraft classification during development[7].

4.2 BAMAS LIBRARY

Unlike the highly controlled setting in which the SERDP library was gathered, the BAMAS library was compiled at each of the three bases from continuously monitored real-world, environmental noise where events exceeding 95 dB were recorded [[17] & [7]].

The BAMAS microphone array, displayed in Figure 13, was developed to reject wind noise and determine the direction of origin of sound using an array of four electret microphones sampled at 5 kHz. Three were parallel to the ground and one was perpendicular. A fifth pre-polarized Type 1 condenser microphone was added to provide higher fidelity measurements for the classifier; which was implemented for blast/non-blast differentiation [10]. All five microphones shared similar sound sensitivities and the same electrical gain was applied to the output signals.

If four of the five microphones exceeded the 95 dB threshold, the BAMAS noise monitors archived the event and the signal in a .csv file. Each .csv file contains the individual microphone thresholds, wind speed, wind heading, time, metrics, and classification results from the UPITT algorithm

Three binary logical checks were used by the BAMAS system to prevent non-acoustic files from being saved. "TDE exceedence" is a binary number linked to the time it takes sound of an event to reach each microphone. If equal to zero, then the time delay between microphones is less than or equal to 9.5 milliseconds; indicating that the event reached all of the microphones at roughly the same time. Should an event cause a time delay greater than 9.5 milliseconds - as is usually the case when wind blows - TDE exceedence would equal one and the event's time history wouldn't be recorded [18].

Figure 13: BAMAS Noise Monitor

Sound clipping indicates whether a microphone exceeded its maximum voltage input of plus or minus 5 V. This would have resulted if either a sound greater than 134 dB was experienced or if the BAMAS noise monitor malfunctioned. Should either of these events occur, the "clipped" logical flag would equal one and the event's time history wouldn't be recorded [18].

Zero phase indicates whether there was any phase (or time delay) between the microphones. Because all blasts have at least some phase, any event without any would be viewed as non-acoustic and excluded from the records [18].

So long as at least four of the five microphones exceeded the 95 dB threshold and none of these logical flags was equal to one, the pressure history from the event was saved. Although several logical flags were used to minimize the recording of non-acoustic events, some wind files still registered as acoustic events. In addition to the acoustic events recorded at Base 3, ten thousand wind files which would have normally been rejected, were recorded and saved as the Wind Library. These files were used to train the ANN so that if future wind files bypass the logical flags, they would be accurately classified as non-acoustic events. Table 4 displays a break down of the number of sound types recorded at each location..

Table 4: File Types from Each Site.

	blast	wind	machine gun	aircraft	vehicle	thunder
Base 1	6,016	385	1,259	448	3,589	221
SERDP	278	566	0	110	0	0
Base 2	3,619	1,737	649	945	1,742	86
Base 4	4,848	129	51	5,501	1,498	3,684
Base 3 (Wind Library)	0	10,103	0	0	0	0
Total	14,761	12,920	1,959	7,004	6,829	3,991

.

4.3 SITE VISIT

A site visit was taken to one of the military bases where BAMAS noise monitors were installed to observe acoustic events in person, verify the acceptance/rejection rates of the original BAMAS noise monitors, and evaluate the accuracy of the original UPITT binary classifier. Files saved from the visit were considered the gold standard because the sound sources were personally verified. Events were observed for two days: one which was exceptionally windy and another which was considerably calmer but was interrupted by a thunderstorm. All human observed events were matched with events logged by the BAMAS noise monitor, but only events which passed the three logical flags were recorded. Figure 14 shows that matched files were located in the dark purple overlap of BAMAS observed and Human observed events. This area included files that were and weren't recorded; however, the light blue area within it contains only files that were observed by both BAMAS and the human *and* were recorded.

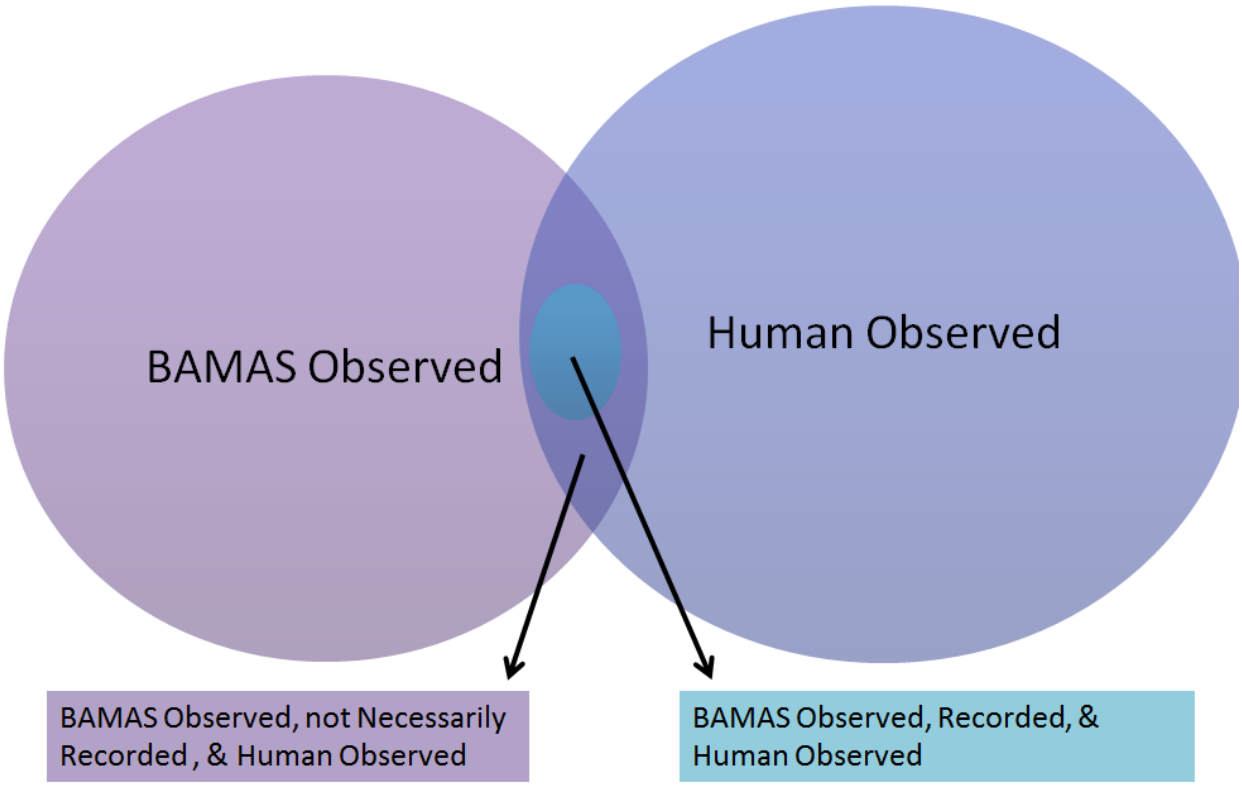

Figure 14: Ven Diagram Differentiating BAMAS Observed and Human Observed

Table 5 contains the rubric used to evaluate accuracy. Figure 15 displays the UPITT algorithm accuracy for the first day - which was windy - while Figure 17 displays the accuracy for the second day - which was calm.

Table 5: Comparison Rubric for BAMAS Observed and Human Observed Files.

True Positive (TP)	UPITT algorithm and Human observer agree sound source was blast
True Negative (TN)	UPITT algorithm and Human observer agree sound source was not a blast
False Negative (FN)	UPITT algorithm said blast event was not a blast event
False Positive (FP)	UPITT algorithm said non-blast event was a blast event

.

Figure 15 shows that the UPITT classifier for cases where files were BAMAS observed, Human observed, but not necessarily recorded has poor blast accuracy during windy conditions. 100% of the blast false positives (FP) and 99% of the false negatives (FN) had raised "TDE exceed" flags due to excessive wind. The overall accuracy (OV AC) was 94.9% because it accurately classified 14,343 wind files which composed 94.7% of the events observed by both the BAMAS and human observers. Although 95.9% accuracy seems high, the problem is that this was designed to be a *blast* classifier. Of the 720 blasts recognized by the BAMAS unit, only 21 were accurately classified representing a scant 3% of the blasts.

Of the 15,140 events logged by the BAMAS noise monitor, only 32 satisfied the wind filtering flags, were recorded, and matched to human observations. 90.6% of those files were accurately classified by the UPITT classifier as shown in Figure 16. Although the overall accuracy of Figure 16, Figure 15 demonstrated how negatively wind can impact the system should it bypass the wind filtering logic.

UPITT classification vs human classification

21	79	2.9%
699	14343	99.5%
79.0%	95.4%	94.9%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

TP w/ tde exceed	TP w/ no raised flags
19.05%	80.95%
FP observed as thunder	FP observed as wind
0.00%	100.00%
FN w/ tde exceed	FN w/ no raised flags
99.14%	0.86%

Figure 15: Original UPITT Binary Classifier vs. Human Observation: Day 1 (BAMAS and HUMAN OBSERVED but not Necessarily Recorded) TP = true positive, FP = False Positive, FN = False Negative, TN = True Negative, OV AC = Overall Accuracy

UPITT classification vs human classification

22	0	88.0%
3	7	100.0%
0.0%	70.0%	90.6%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Figure 16: Original UPITT Binary Classifier vs. Human Observation: Day 1 (BAMAS and HUMAN OBSERVED and Recorded) TP = true positive, FP = False Positive, FN = False Negative, TN = True Negative, OV AC = Overall Accuracy

Figure 17 shows that the UPITT classifier for cases where files were BAMAS observed, Human observed, but not necessarily recorded had 75.5% overall accuracy. Although the blast true positive accuracy was 87%, only 60.9% of the non-blasts were accurately classified. Of the false positives, 26.98% were thunder and 71.43% were wind. Clearly the thunder storm on the second day degraded the accuracy of the classifier in that 94.9% of the false positive were wind induced.

Of the 364 events logged by the BAMAS noise monitor, 285 were saved and matched to human observations; 90.5 % of which were accurately classified by the UPITT classifier as shown in Figure 18. On a clear day, the original UPitt classifier was shown to be highly accurate, but still susceptible to false positives induced by thunder and wind.

UPITT classification vs human classification

177	63	87.2%
26	98	60.9%
26.3%	79.0%	75.5%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

TP w/ tde exceed	TP w/ no raised flags
5.08%	94.92%
FP observed as thunder	FP observed as wind
26.98%	71.43%
FN w/ tde exceed	FN w/ no raised flags
3.85%	96.15%

Figure 17: Original UPITT Binary Classifier vs. Human Observation: Day 2 (BAMAS and HUMAN OBSERVED but not Necessarily Recorded). TP = true positive, FP = False Positive, FN = False Negative, TN = True Negative, OV AC = Overall Accuracy

UPITT classification vs human classification

167	16	93.8%
11	91	85.0%
8.7%	89.2%	90.5%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Figure 18: Original UPITT Binary Classifier vs. Human Observation: Day 2 (BAMAS and HUMAN OBSERVED and Recorded) TP = true positive, FP = False Positive, FN = False Negative, TN = True Negative, OV AC = Overall Accuracy

In summary, the first day of the site visit showed that the UPITT classifier was very adept at accurately classifying wind, but was less successful at blast classification in windy conditions. The second day which was significantly less windy showed much higher blast true positive accuracy (87.2%); however there were 45 wind and 18 thunder induced false positives. In both cases, lower wind sensitivity would have lead to increased overall accuracy.

4.4 HUMAN CLASSIFICATION

Once the sound files from the three bases were collected, each file was listened to, had it's time history and spectra from all microphones plotted, and then classified according to its sound source. Sound files were classified by three different people - the principle graduate student and two assistants. Of the 47,464 files saved, only 11,374 were considered "gold standard". The 10,103 wind library files were verified by APS to be purely wind. The 954 SERDP library files were gathered during the original UPitt classifier's development as discussed earlier. The remaining 317 files were observed and recorded from the Base 3 during the site visit; 203 blasts, 8 thunder, and 106 wind. The principle graduate student did a blind classification test of the site visit data and accurately classified 99.7% of the data, where one mixed blast was inaccurately classified as wind.

A total of 54,783 files were available from all three bases. 11,373 were considered "gold standard" data which were gathered from the SERDP library, wind library, and Base 3 site visit. When classifying the remaining files, the principle graduate only gave classifications to files which were clearly audible as blast, wind, machine gun, vehicle, aircraft, or thunder. Any file which he felt even remotely uncertain about were logged as "unknown" and excluded from training. This process excluded 7,319 unknown files.

After unknowns were excluded, two assistants split the remaining files among themselves and classified them to the best of their abilities. Table 6 shows the consensus between the principle graduate student and his assistants was at least 95% for every sound class. Although there is less certainty in the sound class of the remaining 36,090 files than there was in the "gold standard" data, the level of certainty is only at most 5% less. The files which differed in classification were then reclassified by the principle graduate student.

"Gold standard" data is difficult to collect because it requires the active presence of the researchers at every acoustic event. During the two days of the site visit, only 317 "gold standard" files were captured. This data set did not include enough thunder, aircraft, vehicle, or machine gun data to accurately train the classifier. During the two years of Bucci's research, only 954 "gold standard" samples were collected from highly controlled blast detonations and chance fly overs of aircraft. In order to train the classifier

to accurately classify files which "gold standard" data wasn't available for, a small amount of certainty needed to be sacrificed. By relying exclusively on human classifications, it is possible that human bias was accidentally programmed into the classifier. Attempts were made to quantify the human classification bias, but finding enough willing volunteers to classify 36,000 sound files was more difficult than obtaining "gold standard" data. In the interest of time, the human bias was assumed to be negligible. An increasing sample size analysis was later performed to ensure than enough samples were used to obtain the least possible error.

Table 6: Consensus Between Graduate Student and his Assistants.

blast	wind	machine gun	aircraft	vehicle	thunder
98%	100%	100%	98%	98%	95%

.

5.0 ACOUSTIC METRICS

Additional signal metrics are required for the expanded noise classifier. Twenty-five metrics were evaluated, eleven in the frequency domain and fourteen in the time domain. For the frequency domain metrics, an 8,192 point fast Fourier transform (FFT) was taken of each waveform [10]. All frequency domain metrics were calculated using the magnitude of the FFT, $|G(\omega)|$. The definition for each evaluated metric is as follows:

5.1 SPECTRAL SLOPE

The spectral slope is the slope of a linear fit of the spectra between 2.5 and 100 Hz [6]. It's value, m, is calculated as

$$\widehat{y} = m \times x + b, \tag{5.1}$$

where $\widehat{y} = \log_{10}(FFT)$ is the base-10 logarithm of the fast Fourier transform (FFT) of the signal and $x = \log_{10}(f)$ is that of the base-10 logarithm of the frequency.

5.2 WEIGHTED SQUARE ERROR (WSE)

The weighted square error is the summation of error between the linear fit and the actual value for each frequency bin and is calculated as

$$WSE = \sum_{i=1}^{41} [y_i - \widehat{y}_i]^2 [f_{i+1} - f_i], \tag{5.2}$$

where y_i is the base 10 logarithm of the FFT at the ith frequency bin, $\widehat{y_i}$ is the estimate of the base 10 logarithm of the FFT at the ith frequency bin from the linear fit, and f_i is the base 10 logarithm of the ith frequency bin[6].

5.3 X/Y/LIN/LOG FFT CENTROID

Wind is a relatively pink noise and thus has energy concentrated at very low frequencies. Large blasts are somewhat higher in frequency content, while small arms and aircraft/vehicle are mid and high frequency. With that in mind, the horizontal and vertical centroids for the FFT were calculated as

$$C_x = \frac{\int xdA}{\int ydx} = \frac{1}{6 \times A} \sum_{i=0}^{N-1} x_i + x_{i+1}(x_i y_{i+1} - x_{i+1} y_i), \tag{5.3}$$

and

$$C_y = \frac{\int ydA}{\int ydx} = \frac{1}{6 \times A} \sum_{i=0}^{N-1} x_i + x_{i+1}(x_i y_{i+1} - x_{i+1} y_i), \tag{5.4}$$

respectively for the linear and log-log FFT plots between the frequencies of 2.5 and 200 Hz, where x is the frequency (Hz), y is the magnitude of the FFT, and A is the area beneath the FFT's magnitude plot. The MATLAB implementation of equations 5.3 and 5.4 are also displayed where "*trapz*" was a MATLAB function which finds the trapezoidal integral of the second element with respect to the first and "$cumtrapz(x, y)$" calculates the area beneath the FFT between 2.5 and 200 Hz. Figure 19 displays the linear FFT plot of a sample impulse signal with its centroid displayed with a red "x".

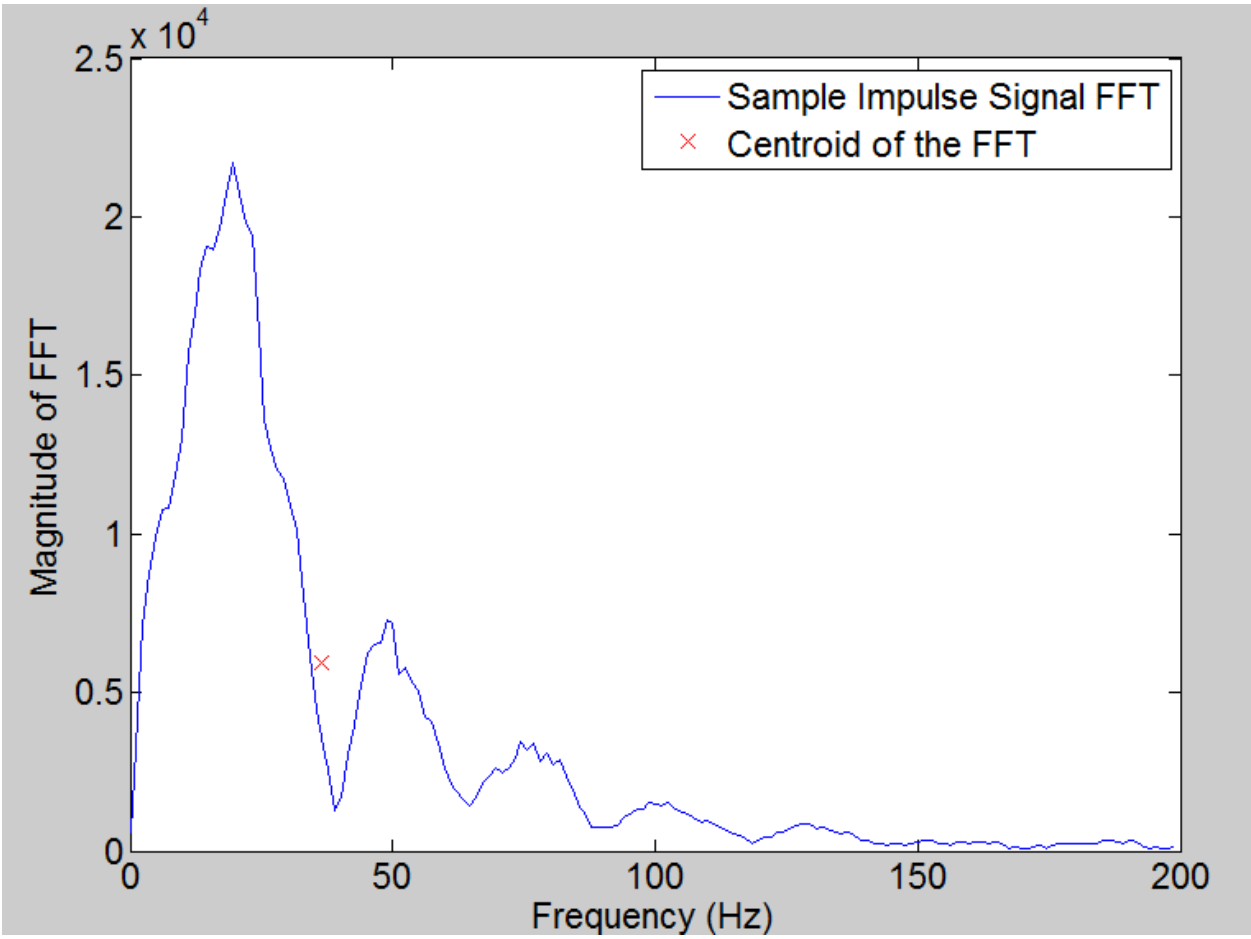

Figure 19: Sample linear FFT with centroid.

5.4 FFT PEAKS

"FFT peaks" refers to the number of peaks in the linear FFT magnitude graph. It was noticed that the FFT magnitude graphs of machine gun and vehicle files had one large peak at the operational frequency of the sound source and multiple harmonics following it. They were calculated using the "peakdet" function in MATLAB[19]. The peakdet function first computes the difference between the mean and the maximum value of a signal vector (δ). Each time the signal passes above δ, a peak is detected.

5.5 KURTOSIS

Kurtosis is the fourth central statistical moment and is calculated as

$$k = \frac{1}{\sigma^4 T} \int_0^T (x - \mu)^4 dt, \tag{5.5}$$

where x refers to the signal, σ is the variance of the signal, μ is the mean acoustic pressure, and T is the time frame over which the kurtosis is measured [6].

5.6 CREST FACTOR

Crest factor is the peak value of the waveform (p_{pk}) divided by the root mean squared value (p_{rms}) of the signal and it is calculated as

$$cf = \frac{p_{pk}}{p_{rms}}. \tag{5.6}$$

5.7 PEAKS

Peaks refers to the number of peaks present in the time history of the waveform. They were also calculated using the peakdet subroutine[19]. During classification, it was noticed that blasts usually had only one or two peaks while machine guns and vehicles had many due to the periodicity with which their sound sources operate.

5.8 A/C/Z FREQUENCY WEIGHTING

Sounds with equal amplitude but different frequency are not perceived by humans to be equally loud. Therefore, to more accurately depict how humans perceive the loudness of a sound in relation to its amplitude, frequency weighting filters were applied to the acoustic signal [20]. Each filter offers a different way of shading a signal to help discern low or high frequency content. The A-weighting filter closely resembles the Fletcher Munson curve at a loudness level of 40 phon[21]. A-weighted sound levels have been shown to correlate well with human response to a variety of environmental noise sources. C-weighting approximates the human ear's response at higher loudness levels (90 phon), and has been shown to correlate better with human response to high-energy impulsive sounds such as blasts. Z-weighting, also known as 'no weighting', is a flat filter and is therefore not a perceptual weighting.

5.9 FAST/SLOW TIME WEIGHTING

Time weighting in the frequency domain of the waveforms is calculated as

$$L_\tau(t) = 10\log_{10}\left(\frac{1}{\tau}\int_{t_s}^{t}\frac{p^2(\zeta)}{p_0^2}e^{\frac{-(t-\zeta)}{\tau}}\right)d\zeta, \tag{5.7}$$

where t_s is the starting time, t is the ending time, p_0 is the reference pressure which equals 20×10^{-6} Pa, p is the pressure (Pa), and ζ is the variable of integration. For fast and slow time weighting, $\tau = 0.125$ and 1 second, respectively [8].

5.10 EQUIVALENT CONTINUOUS SOUND PRESSURE LEVEL

L_{EQ}, the equivalent continuous linearly weighted sound pressure in reference to an atmospheric pressure of 20×10^{-6} Pa over a set time interval,T, [20], is calculated as

$$L_{EQ} = 10\log_{10}\left(\frac{E}{p_0^2T}\right), \tag{5.8}$$

where E refers to the sound exposure which is calculated as

$$E = \int_{t_1}^{t_2} p^2(t)dt, \tag{5.9}$$

and

$$T = t_2 - t_1. \tag{5.10}$$

[8]

5.11 SOUND EXPOSURE LEVEL (SEL)

Sound exposure level is the L_{EQ} normalized to 1 second [20], and thus is calculated as

$$SEL = 10\log_{10}\left(\frac{E}{p_0^2}\right), \tag{5.11}$$

where t_0 (the time length the L_{EQ} is normalized to) is equal to 1 second [8]

5.12 MAX

Max refers to the largest sound level reading in the frequency weighted (A, C, Z) and time-weighted (fast or slow averaging) waveform.

5.13 PEAK

The peak value, "L_{pk}", refers to the maximum frequency-weighted (A, C, Z) waveform and was calculated as

$$L_{pk} = 20\log_{10}\left(\frac{Max\left[p_{A,C,Z}(t)\right]}{2\times 10^{-5}}\right), \tag{5.12}$$

where $p_{A,C,Z}$ refers to the A,C, or Z weighted sound pressure history [8].

6.0 ANN TRAINING

The goal of this study was to train an ANN capable of 90% overall accuracy across six different sound classes. To achieve this, several ANN training techniques were examined and will be discussed in this section.

Proper ANN training aims to find the highest possible classification accuracy while maintaining the generality of the classifier. Sometimes when an ANN is repeatedly trained, it"memorizes" the pattern of the training data set. When evaluating the ANN's efficacy on that data set, the accuracy would be very high; however, if any new data was introduced to the ANN for classification, the accuracy would fall dramatically. Such ANN's are considered "over-trained" and lack the generality required of an effective classifier. Several ANN training techniques exist to maximize classification accuracy; each requiring different amounts of time or number of samples, and employing different evaluation criteria to discourage over-training. Two of those methods were compared in this study: Early Stop and Regularization.

6.1 EARLY STOP METHOD

When using the Early Stop method, the data set is randomly divided into three sections: 60% is set aside for training, 20% for validation, and the remainder is used for testing. Inputs from the training data are fed through the ANN and its outputs are compared to the corresponding targets. The ANN's training error is used to adjust the weights and biases of each neuron to reduce error in future iterations. Next, the validation data - which isn't involved in training - is run through the ANN, the outputs are compared to the targets, and the validation errors are summed. Weights and biases are continuously adjusted until the validation error reaches its minimum and fails to decrease for fifty iterations. This ensures that the network doesn't become "over-trained"; making it useful for only one particular data set. Lastly, the test data is run through the ANN to calculate the network's accuracy. This data set offers the best means of determining the network's accuracy because it wasn't involved in either of the previous steps and demonstrates how the ANN can be expected to perform with new data sets. [22]

6.2 REGULARIZATION METHOD

Regularization prevents over fitting and reduces over complexity by decaying the weights of unnecessary neurons. The method uses all available data to train a network and produce weights and biases. A regularization coefficient (λ) is used in conjunction with the mean squared error (mse) to produce the regularized mean square error as shown in equation (6.1)

$$mse_{reg} = \lambda \times mse + (1 - \lambda) \times msw, \tag{6.1}$$

where msw is the mean square of the weights and biases. The changes to the weights and biases are calculated using equation (6.2)

$$dW = -(jj + I \times \mu) \times [je]^{-1}, \tag{6.2}$$

where jj is the Jacobian of the weights and biases, I is an identity matrix, je is the Jacobian of the regularized mean square errors, and μ is a number systematically adjusted to produce the lowest mean square error. Training concludes either when all possible values of μ have been used or the mean square error reaches zero. The advantage of this method over the "Early Stop" method is that it requires less samples. Unfortunately, because it varies μ and λ, it requires significantly more time for training than the early stop method [22].

Two ANN's were calculated using both methods. Because the difference between the classification accuracy was minimal and the training time for the Regularization method was significantly longer than that of the early stop, the later was chosen as the ANN training technique to be utilize during this study.

7.0 ANN EVALUATION

7.1 ROUND METHOD VERSUS MAX METHOD

Output vectors of an ANN are 6×1 vectors whose values can be any number due to the purelin activation function of the output layer; however output values are heavily focused between 0 and 1 due to the logsig transfer function of the third hidden layer. Target vectors are also 6×1 vectors populated by five 0's and a 1 whose position indicates the source of the noise source. To get the output vectors to match the target vectors, two methods were examined. The "round" method rounded output vectors greater than .5 to 1 and other values are rounded to 0. It was utilized in the original UPITT algorithm; however, with multiple classes, there is the risk that multiple outputs could round up to 1; indicating multiple sound sources. Conversely, all of the outputs could round down to 0 which would result in a null classification. The "max" method makes the largest output vector equal to 1 and the rest equal to 0. By examining only the largest ANN output, this method avoids the possibility of multiple or null classifications.

7.2 CONFUSION MATRICES

Once the network was trained, all of the data was run through the ANN. Since there were 6 possible outputs: blast, wind, aircraft, vehicle, machine gun, thunder, there were six ANN outputs with values of either 0 or 1. A value of 1 signifies the ANN's prediction for the sound source. Table 7 shows how the predictions of the ANN were compared with the targets.

Table 7: Comparison Rubric.

True Positive (TP)	Both the target and output are equal to one
True Negative (TN)	Both the target and output are equal to zero
False Negative (FN)	The target is equal to one, but the output is equal to zero
False Positive (FP)	The target is equal to zero, but the output is equal to one

.

Figure 20 displays a 7×7 confusion matrix of an ANN trained using the four original metrics and evaluated using the max method. These matrices allowed researchers to find which sound classes were most likely to be confused for each other. The columns signify the targets and the rows signify the ANN's predictions. The green numbers along the diagonal signify when the ANN's predictions match the targets' also known as true positives (TP). The blue number was the overall accuracy: the sum of true positives divided by the total number of instances. The somewhat low accuracy (74.35%) demonstrated that the classifier needed improvement in terms of finding new metrics which will be the focus of this study.

Each of the red numbers were incorrect classifications. False positives populated the upper diagonal. The number in first row second column, 173, was the number of wind files mistaken as blasts. Similarly 1,301 machine gun files were falsely classified as blasts. False negatives populated the lower diagonal. The number in the second row first column, 515, was the number of blast files mistaken as wind. Likewise 141 blasts were falsely classified as machine guns. In the gray column to the right were the percentages of specific noise types marked as false positives (e.g. 16.81% of the files labeled as blasts were false positives while 12.36% of files marked as wind were false positives). The gray row at the bottom contains the false negative percentages (e.g. 13.31% of blasts were incorrectly labeled as another sound class, while 91.86% of machine gun files were incorrectly labeled as another sound class). It should be noted that the 6×6 ANN doesn't have true negatives in the traditional sense. Every true positive classification for one class is a true negative for every other class. When calculating the overall accuracy, the total correct classifications (i.e. true positive/true negative) is divided by the total number of instances.

Figure 20: Confusion matrix: all (original four metrics and max method.)

.

The six 3×3 confusion matrices in Figure 20 are the binary confusion matrices for each noise type. Notice how the number in the first row first column of the blast 3x3 confusion matrix (12,797) was the same as that of the first row first column of the 7×7 confusion matrix. Each individual confusion matrix breaks down the effectiveness of the ANN to properly classify each noise type individually. Blast/non-blast classification takes priority over the other sound types because blast identification was the focus of this work; however, it was nice to know how well the classifier was able to classify other sound types individually.

7.3 ORIGINAL METRIC ANN ACCURACY

Figure 21 displays the accuracy of an ANN trained using the "Early Stop Method", only the original four metrics (kurtosis, weighted square error, spectral slope, and crest factor), and evaluated using the "round" method discussed in Section 6.3.1. It is similar to the ANN computed in Figure 20, except that the later was evaluated using the "Max" method discussed in Section 6.3.1. Under controlled circumstances, blast, wind and aircraft classification accuracy using only those metrics had around 99% accuracy[6]; however, when samples from several different locations were classified using those metrics, blast, wind, and aircraft classification accuracy diminished by an average of 8%. Machine Gun and thunder accounted for 87% of the blast false positives. There were incidents when the ANN's predictions rounded to zero; resulting in 10,852 null classification which are displayed in Table 8.

Predictions \ Targets	blast	wind	mach	air craft	vehicle	thunder	
blast	11718	86	739	20	87	512	10.97%
wind	386	11737	52	350	286	79	8.94%
mach	0	0	0	0	0	0	0.00%
air craft	177	50	62	3348	1561	216	38.16%
vehicle	37	32	4	523	2526	35	19.99%
thunder	306	33	50	184	162	1357	35.13%
	7.18%	1.68%	0.00%	24.34%	45.35%	38.29%	64.51%

Blast

11718	1444	79.38%
3043	31362	95.60%
4.40%	20.62%	90.57%

Air Craft

3348	2066	47.11%
3759	38394	94.89%
5.11%	52.89%	87.75%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Wind

11737	1153	90.84%
1183	33494	96.67%
3.33%	9.16%	95.09%

Vehicle

2526	631	36.99%
4303	40107	98.45%
1.55%	63.01%	89.63%

Machine Gun

0	0	0.00%
1959	45608	100.00%
0.00%	100.00%	95.88%

Thunder

1357	735	34.00%
2634	42841	98.31%
1.69%	66.00%	92.92%

Figure 21: Original Four Metric ANN Confusion Matrix (Round Method)

Table 8: Round Method Null Classifications (Original Four Metrics).

Blast	2137
Wind	982
Machine Gun	1052
Air craft	2682
Vehicle	2207
Thunder	1792

.

Summary: Overall accuracy of the ANN trained with the original four metrics for blast, wind, and aircraft was significantly lower for the combined SERDP and BAMAS library data set.

7.4 NEW METRIC ANN ACCURACY

When all twenty five metrics were used, overall accuracy was increased by 11.36% as shown in Figure 22. There were still an undesirable number of null classifications as shown in Table 9.

Predictions \ Targets	blast	wind	mach	air craft	vehicle	thunder	
	13929	112	105	10	20	401	4.45%
	112	12223	6	51	152	33	2.81%
	143	6	1465	10	52	25	13.87%
	7	53	17	5514	1054	39	17.50%
	28	114	44	653	4526	71	16.74%
	251	44	12	103	57	3112	13.05%
	3.74%	2.62%	11.16%	13.04%	22.78%	15.46%	85.71%

Blast		
13929	648	94.36%
832	32158	98.02%
1.98%	5.64%	96.89%

Air Craft		
5514	1170	77.59%
1593	39290	97.11%
2.89%	22.41%	94.19%

Rubric		
#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Wind		
12223	354	94.61%
697	34293	98.98%
1.02%	5.39%	97.79%

Vehicle		
4526	910	66.28%
2303	39828	97.77%
2.23%	33.72%	93.25%

Machine Gun		
1465	236	74.78%
494	45372	99.48%
0.52%	25.22%	98.47%

Thunder		
3112	467	77.98%
879	43109	98.93%
1.07%	22.02%	97.17%

Figure 22: ALL Metrics ANN Confusion Matrix

Summary: Then new metric boosted overall accuracy by 11.36%, but the round method still produces null classifications.

Table 9: Round Method Null Classifications (All Metrics (Round Method)).

Blast	291
Wind	368
Machine Gun	310
Air craft	968
Vehicle	766
Thunder	310

7.5 ROUND VS. MAX METHOD

The 10.17% increase in overall accuracy from Figure 20 to 21 shows that the max method is more effective. This is further emphasized by the 1.83% increase in overall accuracy from Figure 22 to 23. Although in the latter case the increase in overall accuracy was small, by using the "Max" method, there were zero null classifications; making it the preferable to the round method.

Summary: Max method increased accuracy of the ANN trained with the new metrics by 10.17% and resulted in zero null classifications. Similarly, the max method increased overall accuracy of the ANN trained using the old metrics by 9.84%. Clearly, the max method is preferable to the round method.

blast	wind	mach	air craft	vehicle	thunder	
14094	153	151	10	48	440	5.38%
159	12351	19	80	180	81	4.03%
160	19	1659	29	67	29	15.49%
12	53	26	5746	952	52	16.01%
42	257	77	1035	5459	149	22.23%
294	87	27	104	123	3240	16.39%
4.52%	4.40%	15.31%	17.96%	20.06%	18.82%	89.64%

Blast

14094	802	95.48%
667	31901	97.55%
2.45%	4.52%	96.91%

Air Craft

5746	1095	82.04%
1258	39365	97.29%
2.71%	17.96%	95.04%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Wind

12351	519	95.60%
569	34025	98.50%
1.50%	4.40%	97.71%

Vehicle

5459	1560	79.94%
1370	39075	96.16%
3.84%	20.06%	93.83%

Machine Gun

1659	304	84.69%
300	45201	99.33%
0.67%	15.31%	98.73%

Thunder

3240	635	81.18%
751	42838	98.54%
1.46%	18.82%	97.08%

Figure 23: ALL Metric ANN Confusion Matrix (Max Method)

7.6 REFINED CLASSIFICATIONS

Although the new overall accuracy was 89%, there was a large amount of confusion between aircraft and vehicles. Additionally there were 440 thunder files mistaken as blasts. Initially the data was classified by three different people; each with their own interpretations of what each sound file sounds like due to their personal experiences. To ensure the homogeneity of the classification criteria, the principle graduate student reclassified the sound files whose classifications differed between the three people; producing the ANN in Figure 24 which shows a 3.26% increase in overall accuracy. Aircraft and vehicle confusion fell by 64%. Additionally, blast and thunder confusion only fell by 12% and overall accuracy surpassed the 90% goal.

blast	wind	mach	air craft	vehicle	thunder	
14159	131	127	3	40	411	4.79%
150	12440	14	28	270	39	3.87%
120	8	1752	17	81	21	12.36%
4	9	21	3135	337	45	11.72%
46	278	63	374	9241	188	9.31%
232	84	23	44	156	3306	14.02%
3.75%	3.94%	12.40%	12.94%	8.73%	17.56%	92.90%

Blast

14159	712	96.25%
552	31974	97.82%
2.18%	3.75%	97.33%

Air Craft

3135	416	87.06%
466	43380	99.05%
0.95%	12.94%	98.14%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Wind

12440	501	96.06%
510	33946	98.55%
1.45%	3.94%	97.87%

Vehicle

9241	949	91.27%
884	36323	97.45%
2.55%	8.73%	96.13%

Machine Gun

1752	247	87.60%
248	45150	99.46%
0.54%	12.40%	98.96%

Thunder

3306	539	82.44%
704	42848	98.76%
1.24%	17.56%	97.38%

Figure 24: ALL Metric ANN Confusion Matrix (Refined Classifications)

Summary: Reclassifying helicopter and small single engine planes as vehicle decreased aircraft/vehicle confusion by 64%.

7.7 MIXED BLASTS

Since mixed blasts are a composition of both blasts and wind, one may ask whether those files should be classified as blasts even if the blast is overpowered by wind. Considering that the goal of the project was to accurately classify blasts, it was imperative that any event containing a blast be classified as such. The question then became whether or not those mixed blasts should be incorporated during training. Two ANN's were trained: one using mixed blasts during training and one that didn't. Select portions of the data were then run through the two ANN's to test their ability to properly classify ideal blasts, mixed blasts, and wind files. Figures 25 and 26 display that in both cases the ANNs' were able to accurately classify ideal blasts with 98% accuracy; however, the inclusion of mixed blasts during training boosted the ANN's ability to properly classify mixed blasts by 30% from 60.62% to 89.92%. Therefore, mixed blasts should be included during training.

Just the Blasts

blast	wind	mach	air craft	vehicle	thunder	
10672	0	0	0	0	0	0.00%
18	0	0	0	0	0	0.00%
118	0	0	0	0	0	0.00%
2	0	0	0	0	0	0.00%
11	0	0	0	0	0	0.00%
109	0	0	0	0	0	0.00%
2.36%	0.00%	0.00%	0.00%	0.00%	0.00%	97.64%

Blast			Rubric		
10672	0	97.64%	#TP	#FP	%TP
258	0	0.00%	#FN	#TN	%TN
0.00%	2.36%	97.64%	%FP	%FN	%OV AC

Just the Mixed Blasts

blast	wind	mach	air craft	vehicle	thunder	
2292	0	0	0	0	0	0.00%
765	0	0	0	0	0	0.00%
101	0	0	0	0	0	0.00%
6	0	0	0	0	0	0.00%
21	0	0	0	0	0	0.00%
596	0	0	0	0	0	0.00%
39.38%	0.00%	0.00%	0.00%	0.00%	0.00%	60.62%

Blast			Rubric		
2292	0	60.62%	#TP	#FP	%TP
1489	0	0.00%	#FN	#TN	%TN
0.00%	39.38%	60.62%	%FP	%FN	%OV AC

Just the Winds

blast	wind	mach	air craft	vehicle	thunder	
0	27	0	0	0	0	0.00%
0	12439	0	0	0	0	0.00%
0	12	0	0	0	0	0.00%
0	9	0	0	0	0	0.00%
0	299	0	0	0	0	0.00%
0	164	0	0	0	0	0.00%
0.00%	3.95%	0.00%	0.00%	0.00%	0.00%	96.05%

Wind			Rubric		
12439	0	96.05%	#TP	#FP	%TP
511	0	0.00%	#FN	#TN	%TN
0.00%	3.95%	96.05%	%FP	%FN	%OV AC

Figure 25: Mixed Blasts Excluded from Training

Just the Blasts

blast	wind	mach	air craft	vehicle	thunder	
10736	0	0	0	0	0	0.00%
15	0	0	0	0	0	0.00%
94	0	0	0	0	0	0.00%
1	0	0	0	0	0	0.00%
15	0	0	0	0	0	0.00%
69	0	0	0	0	0	0.00%
1.77%	0.00%	0.00%	0.00%	0.00%	0.00%	98.23%

Blast		
10736	0	98.23%
194	0	0.00%
0.00%	1.77%	98.23%

Rubric		
#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Just the Mixed Blasts

blast	wind	mach	air craft	vehicle	thunder	
3400	0	0	0	0	0	0.00%
132	0	0	0	0	0	0.00%
39	0	0	0	0	0	0.00%
3	0	0	0	0	0	0.00%
15	0	0	0	0	0	0.00%
192	0	0	0	0	0	0.00%
10.08%	0.00%	0.00%	0.00%	0.00%	0.00%	89.92%

Blast		
3400	0	89.92%
381	0	0.00%
0.00%	10.08%	89.92%

Rubric		
#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Just Wind

blast	wind	mach	air craft	vehicle	thunder	
0	152	0	0	0	0	0.00%
0	12364	0	0	0	0	0.00%
0	16	0	0	0	0	0.00%
0	18	0	0	0	0	0.00%
0	272	0	0	0	0	0.00%
0	128	0	0	0	0	0.00%
0.00%	4.53%	0.00%	0.00%	0.00%	0.00%	95.47%

Wind		
12364	0	95.47%
586	0	0.00%
0.00%	4.53%	95.47%

Rubric		
#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Figure 26: Mixed Blasts Included in Training

Summary: Excluding mixed blasts from classification hindered mixed blast classification. The original UPitt classifier was not trained with mixed blasts. This study indicates that if it had, the classifier may have been better able to accurately classify mixed blasts.

7.8 INCREASING SAMPLE SIZE ANALYSIS (ISA)

When training an ANN, it's important to know how many samples is enough. Too few, and there won't be enough samples to properly train the network, but too many will unnecessarily increase training time. To determine the critical number of samples, five ANN's were trained with one thousand samples randomly selected from the data set. Of the five ANN's, the one with the highest test accuracy was selected to represent an ANN trained using one thousand samples. The process was then repeated using two thousand randomly selected samples, then three thousand, and continued until all available samples were used Once finished, the training, validation, and test errors were plotted as a function of number of samples used in training.

Increasing sample size analysis (ISA) plots, shown in Figures 27 through 29, provide several beneficial pieces of information. Each figure uses data from one of three sites. The critical number of samples occurs where the test error stops decreasing. If at any point the test error is much higher than the validation or training error, it means that the ANN has poor generality. Should the validation and test errors decline at roughly the same rate as the number of samples used in training increases, it means the data has been properly divided throughout the three sections.

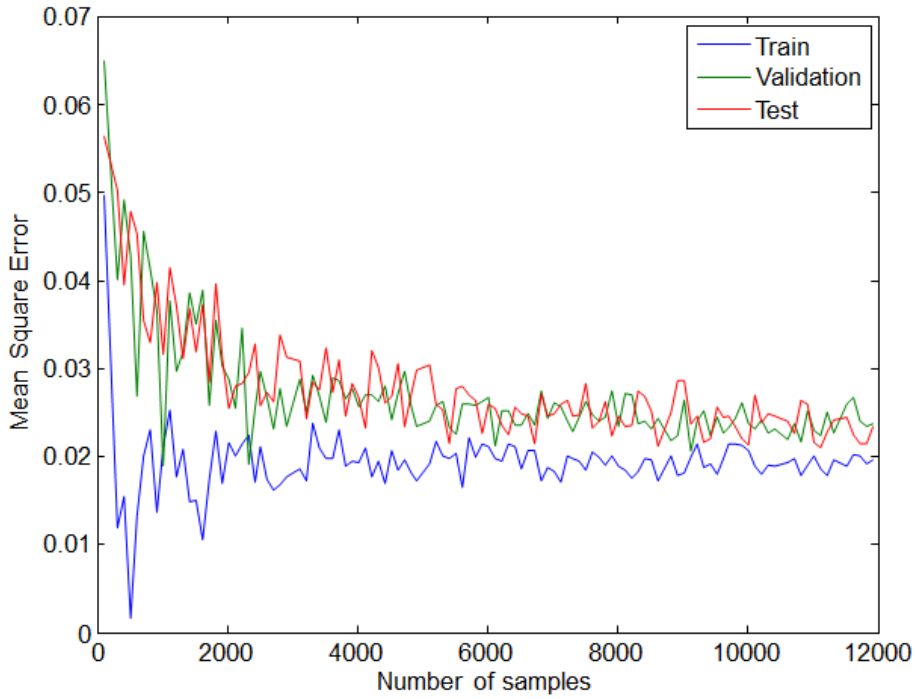

Figure 27: Increasing Sample Size Analysis Results for Base 1

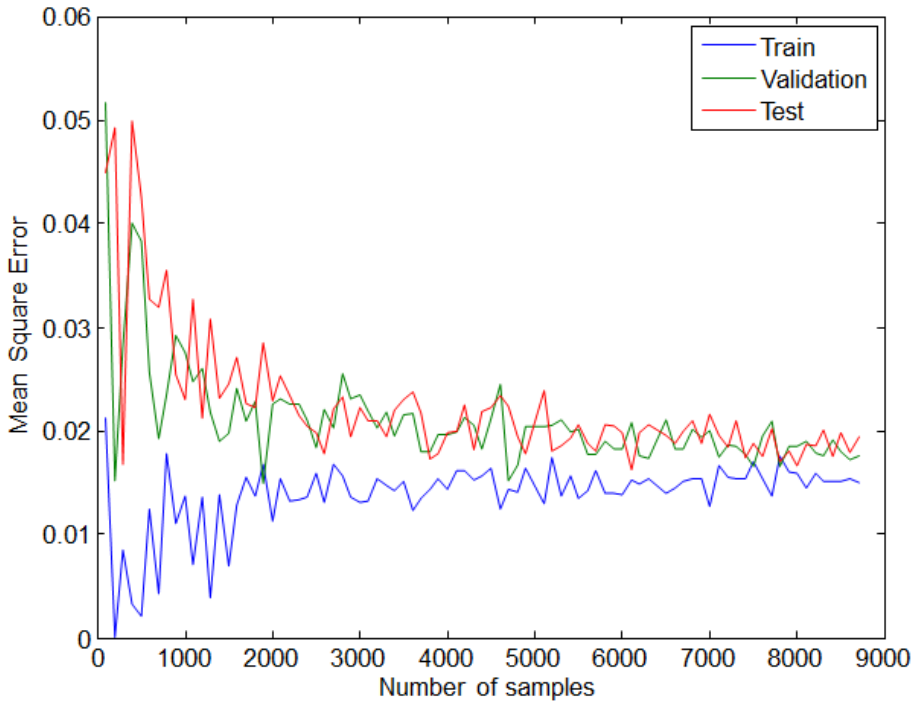

Figure 28: Increasing Sample Size Analysis Results for Base 2

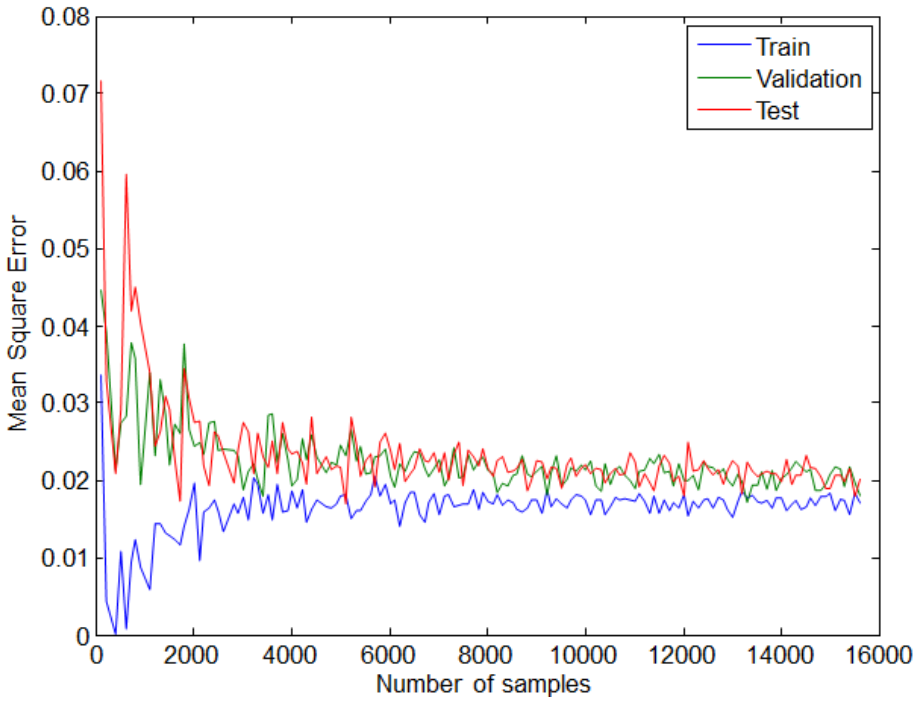

Figure 29: Increasing Sample Size Analysis Results for Base 3

According to Figures 27- 29, ANN's trained and tested with data from the site from which the data was gathered closely approaches the minimum mean square error after about three to four thousand samples. Figure 30 shows that when all of the data is lumped together, mean square error decelerates after about thirty four thousand samples. Because more samples were used for training than were necessary to produce the minimum mean square error, these graphs imply that enough samples were being used to yield the maximum possible accuracy for each site individually as well as for the general classifier.

In Figure 30, validation and testing errors declined at roughly the same rate indicating proper data division. The two errors also remained roughly the same as the number of samples increased - indicating good ANN generality. Therefore, any ANN trained with a data set larger that 34,000 samples will be able to effectively train the ANN. It should be noted that the minimum mean square error for the "All" data set was higher than that of the others indicating that site specific training should yield increased accuracy; however, due to the intended nature of the final device such training was infeasible. Most who would purchase a sound classification system expect maximum accuracy immediately once it is installed. Site specific training would require at least a month to gather sound samples, human classify them, train an ANN, and upgrade the system.

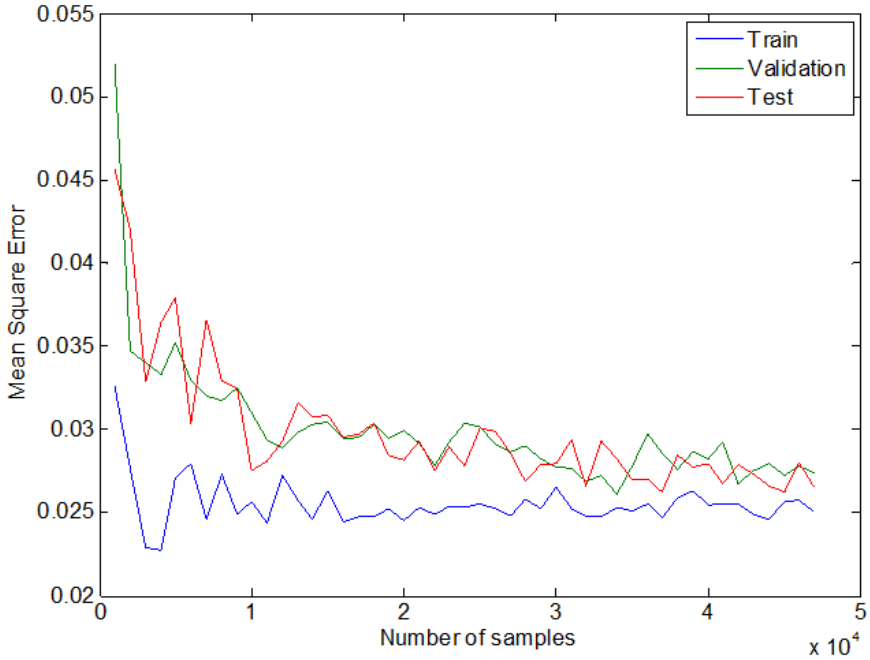

Figure 30: Increasing Sample Size Analysis Results for All Data

7.9 REMOVAL OF WIND FROM TRAINING AND EVALUATION

The three flag wind filtering technique discussed earlier was effective at rejecting wind noise, but not perfect. The question then arose as to the efficacy of including wind files in training if wind files were routinely rejected before being saved. More importantly, what would happen should wind files bypass the filters. An ANN was trained without using any wind files and evaluated using wind and non-wind files. Figures 31 - 32 show that exclusion of wind files from training has minimal effect on the classification of other file types; however, should wind files bypass the filters, they would most likely be incorrectly classified as blast or vehicle false positives. Considering that the purpose of the device is to properly classify blast noise, having a system that mistakes a common occurrence as blast was particularly problematic. For this reason, it was recommended that wind files be included during the training process to err on the side of caution.

blast	wind	mach	air craft	vehicle	thunder	
14293	8019	137	6	82	360	37.58%
0	0	0	0	0	0	0.00%
107	79	1741	13	56	13	13.34%
9	77	19	3148	312	35	12.56%
49	4544	80	494	9527	220	36.12%
253	231	23	43	148	3382	17.11%
2.84%	0.00%	12.95%	15.01%	5.91%	15.66%	67.56%

Blast

14293	8604	97.16%
418	24185	73.76%
26.24%	2.84%	81.01%

Air Craft

3148	452	84.99%
556	43344	98.97%
1.03%	15.01%	97.88%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Wind

0	0	0.00%
12950	34550	100.00%
0.00%	100.00%	72.74%

Vehicle

9527	5387	94.09%
598	31988	85.59%
14.41%	5.91%	87.40%

Machine Gun

1741	268	87.05%
259	45232	99.41%
0.59%	12.95%	98.89%

Thunder

3382	698	84.34%
628	42792	98.40%
1.60%	15.66%	97.21%

Figure 31: ANN Trained Using no Wind Data During Training, but Wind Files are Included in the Evaluation

blast	wind	mach	air craft	vehicle	thunder	
14293	0	137	6	82	360	3.93%
0	0	0	0	0	0	0.00%
107	0	1741	13	56	13	9.79%
9	0	19	3148	312	35	10.64%
49	0	80	494	9527	220	8.13%
253	0	23	43	148	3382	12.13%
2.84%	0.00%	12.95%	15.01%	5.91%	15.66%	92.88%

Blast

14293	585	97.16%
418	19254	97.05%
2.95%	2.84%	97.10%

Air Craft

3148	375	84.99%
556	30471	98.78%
1.22%	15.01%	97.31%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Wind

0	0	0.00%
0	34550	100.00%
0.00%	0.00%	100.00%

Vehicle

9527	843	94.09%
598	23582	96.55%
3.45%	5.91%	95.83%

Machine Gun

1741	189	87.05%
259	32361	99.42%
0.58%	12.95%	98.70%

Thunder

3382	467	84.34%
628	30073	98.47%
1.53%	15.66%	96.83%

Figure 32: ANN Trained Using no Wind Data During Training or Evaluation

Summary: Although wind isn't expected to bypass the wind filter logic, if it does it would most likely become a blast false positive if wind isn't included during training.

7.10 SIGNAL TO NOISE RATIO (SNR)

Figure 20 displayed that 515 blast files were mistaken as wind by the classifier. Upon further investigation it was determined that these sound samples contained blasts but were heavily contaminated by wind. To determine the ratio of blast signal to wind signal that would result in an improper classification, a signal to noise ratio analysis was performed where "signal" referred to pure blast samples and "noise" exclusively referred to wind samples. Equation 7.1 defines the SNR,

$$SNR_{desired(dB)} = 10 \times \log_{10}\left(\frac{K^2 \times P^2_{blast}}{P^2_{wind}}\right), \tag{7.1}$$

where "P^2_{signal}" was the mean square amplitude of a randomly chosen signal ($P^2_{signal} = \frac{1}{N}\sum_{i=1}^{N} signal_i^2$), and K was a constant chosen to force the SNR of the synthetically fabricated mixed blast sound pressure to conform to desired SNR values. K was calculated using equation 7.2 for each desired SNR,

$$K = \sqrt{\frac{P^2_{wind} \times 10^{(SNR_{desired}/10)}}{P^2_{blast}}}. \tag{7.2}$$

Equation 7.3 combined 1000 randomly chosen blast and wind samples into synthetically fabricated mixed blast samples,

$$P_{mixed} = P_{blast(i,:)} + \frac{1}{K} \times P_{wind(i,:)}, \tag{7.3}$$

where i cycles through 1000 blast and 1000 wind time histories and K changes depending on the desired SNR. Each of the 1000 blast time histories was matched with the same wind time history every time. The product of Equation 7.3 was 1,000 mixed blast samples for each desired SNR. Ideally every file would be classified as a blast for each desired SNR; however, those that didn't would demonstrate how sensitive the ANN classification system was to wind noise.

Figure 33 shows that at an SNR of 40 dB, the classifier accurately identifies blasts 99.9 % of the time. Correct blast classification fell to 99% at blast to wind ratios of roughly 10 dB. Considering that the ANN was able to correctly classify blasts 96% of the time and the SNR shows a blast accuracy of 95.5% at an SNR of 0 dB, it can be assumed that the trained ANN is moderately insensitive to wind under typical conditions. Should the wind be stronger that usual (ie. in a severe storm) the classifier can be expected to give many blast false negatives. Otherwise blast true positive accuracy should be expected to remain high.

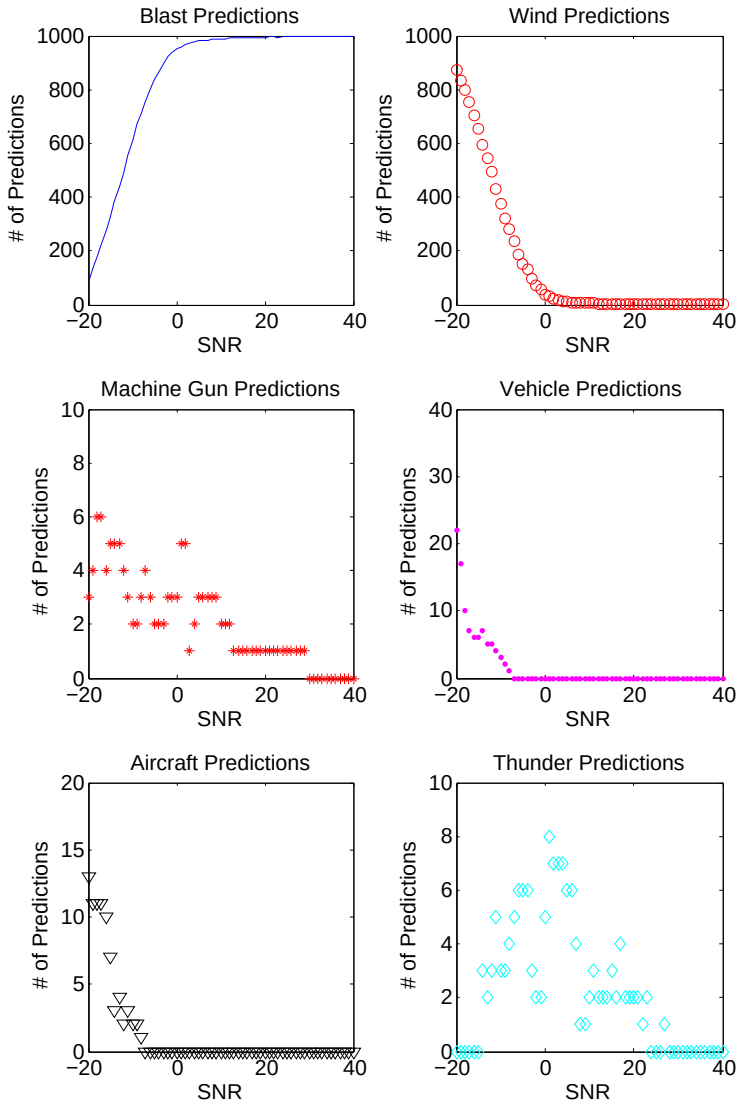

Figure 33: Classifier Performance at Various Signal (pure blast) to Noise (pure wind) Ratios

7.11 HUMAN CLASSIFICATION COMPARISON

The goal of this work was to train an ANN that could accurately classify new sound samples; however, there are limits to how accurate it can reasonably be expected to perform. Before training could begin, thousands of sound samples were collected and listened to by human beings. Like the ANN, people's brains are trained to identify noise using personal experience, sound, and context clues from the environment. Unfortunately all sound samples were recorded as .bin files; robbing those who need to classify them of environmental context clues. Determining aircraft from blast may be simple if only given the sound, but differentiating blast from thunder without knowledge of the weather is especially challenging. Similarly, not knowing what machine guns sound like or knowing to expect them would make small blast (22 mm ordinance) and machine gun classification significantly more difficult.

To determine the variation of human classification of sound files while making the task fair, a training regimen was designed to teach anyone how to classify sounds using a program written for this study. Seven test subjects (four engineers and three non-engineers) were shown the same Power Point presentation detailing what each sound file's wave form looks like in graphical form (as seen in Figures 4 - 11), and given 1001 samples to classify. The 1001 samples contained 286 blasts/mixed blasts, and 143 of everything else. Their accuracies were then compared with that of the ANN.

Figure 34 shows boxplots with the human accuracies along with the ANN accuracy shown as a circle. It is clear when comparing the median accuracy of the humans to the ANN accuracy, the new ANN was more accurate than most humans at correctly classifying blast true positives and false positives, but lacked in its ability to discern true negatives and false negatives. Its overall blast accuracy, from the last column of Figure 34, was comparable, but slightly less than, that of the test subjects.

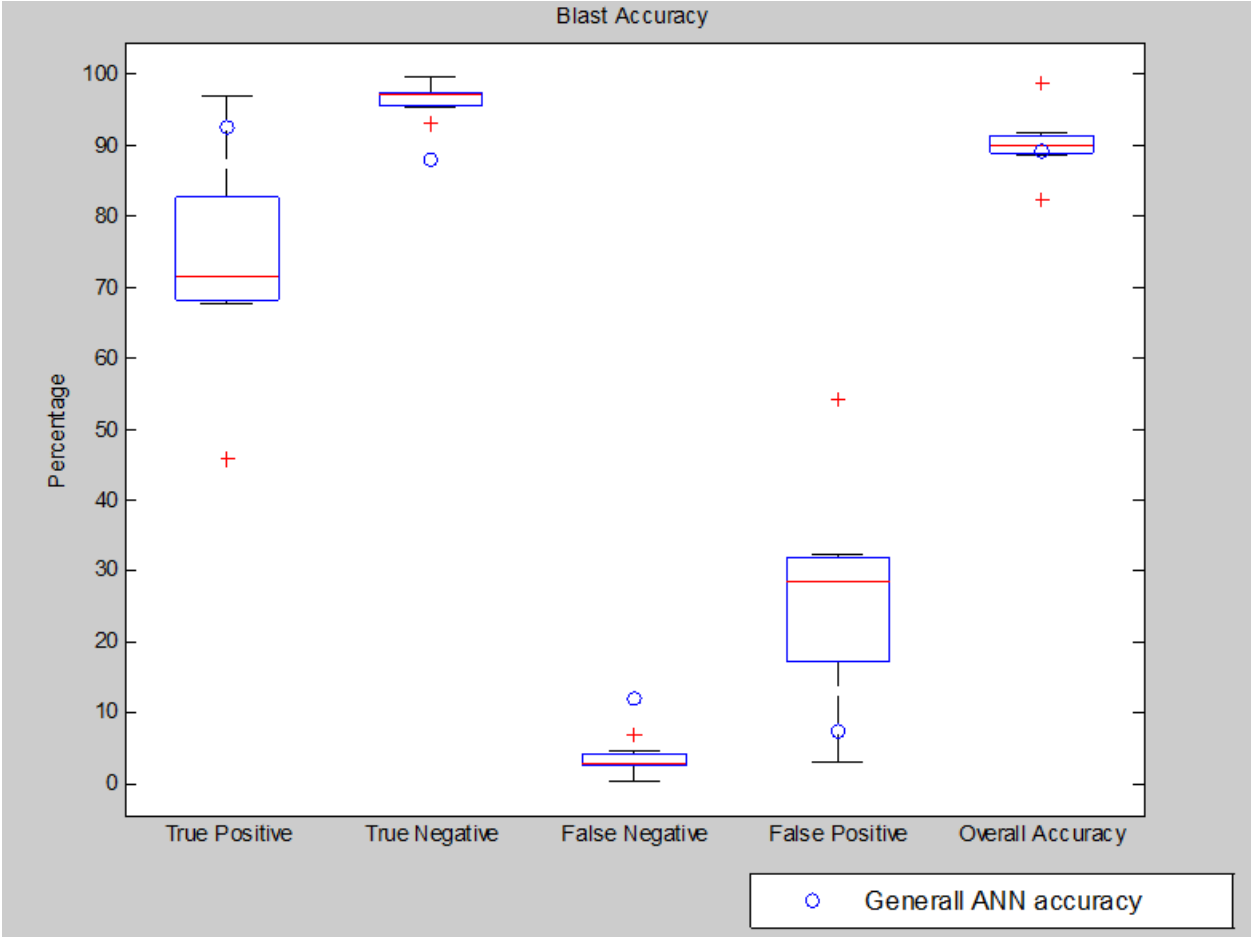

Figure 34: Human to ANN Accuracy Comparison: Blast

According to Figures 35 and 36, the ANN was significantly less accurate when identifying wind and machine gun true positives and false positives, however it performed better than the test subjects at discerning true negatives and false negatives. Wind overall accuracy fell in line with that of the test subjects, but machine gun overall accuracy fell below the 25th percentile of human accuracies; indicating it's inferior performance.

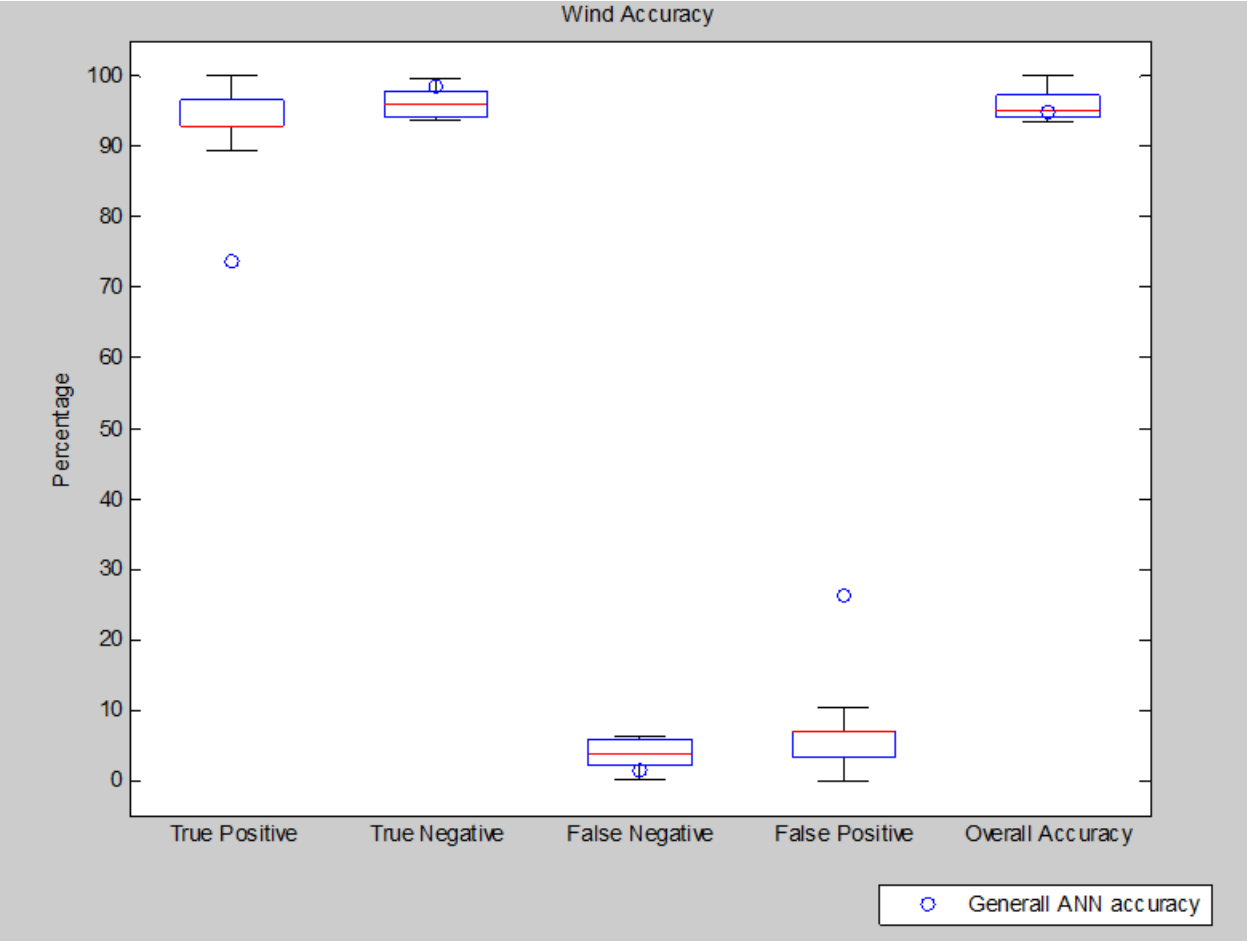

Figure 35: Human to ANN Accuracy Comparison: Wind

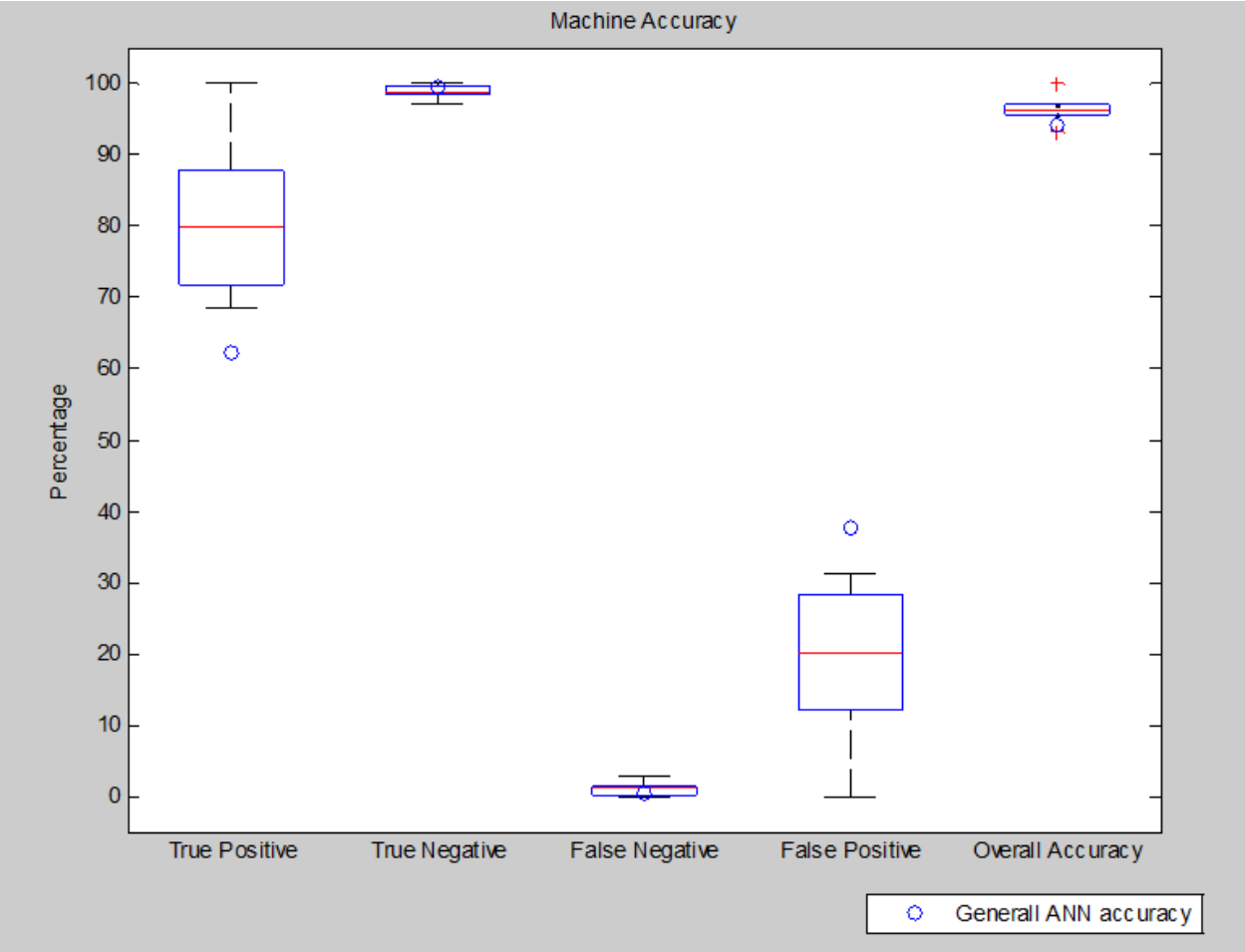

Figure 36: Human to ANN Accuracy Comparison: Machine Gun

Figures 37 and 38 show that the ANN's aircraft and vehicle accuracies surpassed the 75th percentile of human accuracies; proving to be slightly better.

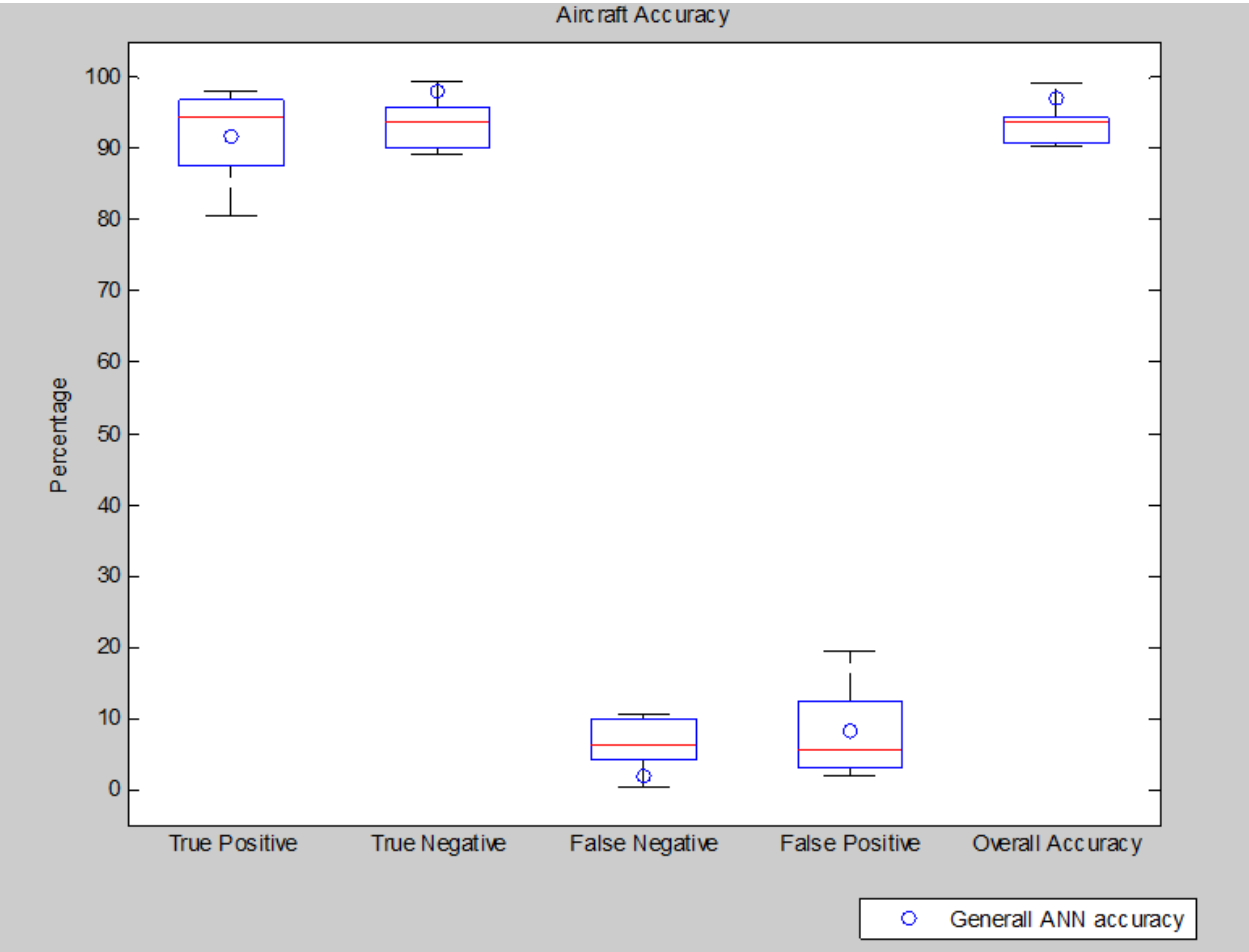

Figure 37: Human to ANN Accuracy Comparison: Aircraft

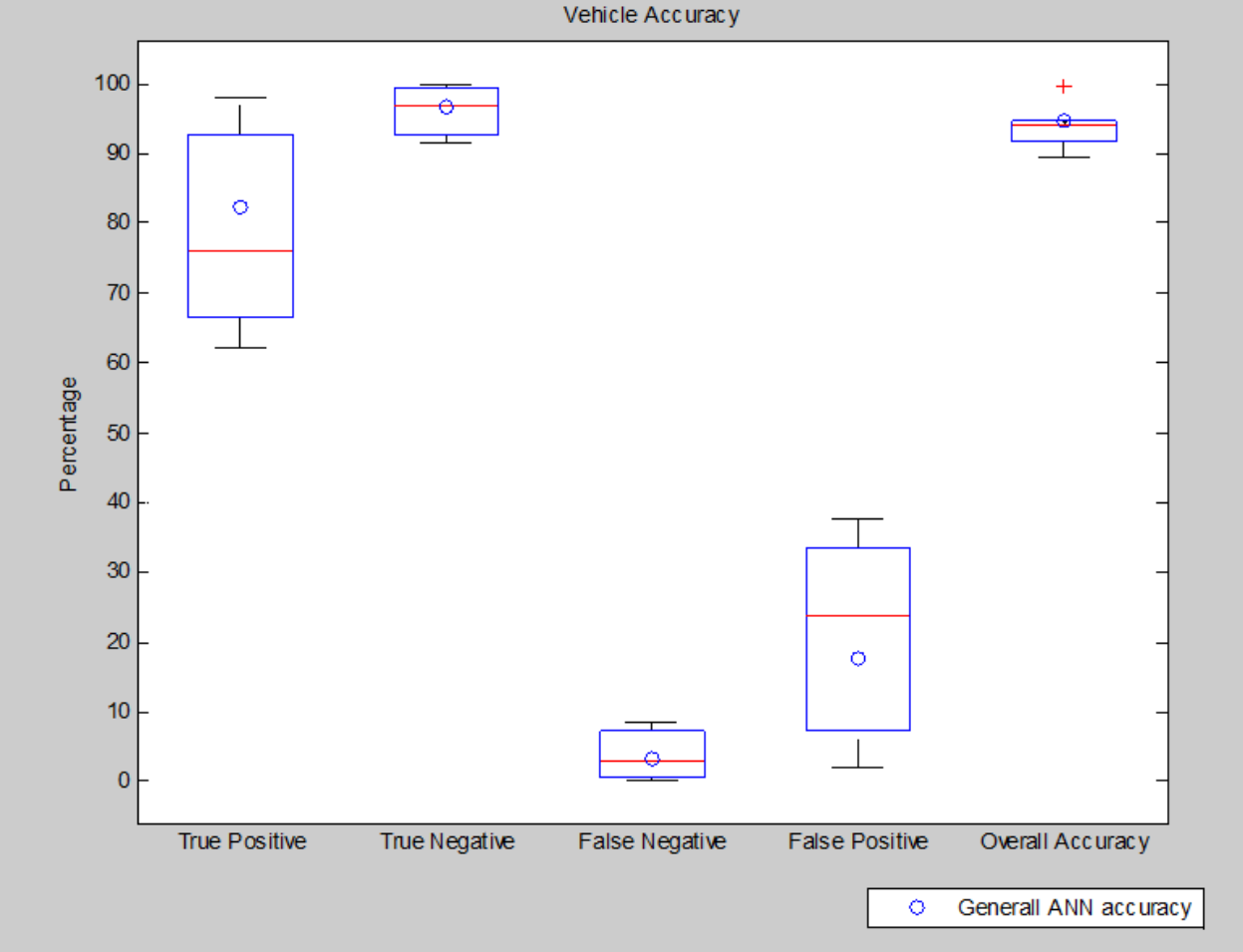

Figure 38: Human to ANN Accuracy Comparison: Vehicle

Figure 39 shows that the ANN was 9% more accurate than the median test subject at correctly identifying thunder true positives. Both were roughly equivalent at discerning thunder true negatives, and false negatives. Although the ANN scored better than the median test subject at classifying thunder false positives, both false positive rates were undesirably high. Overall, the ANN classified thunder just as effectively as the median test subject.

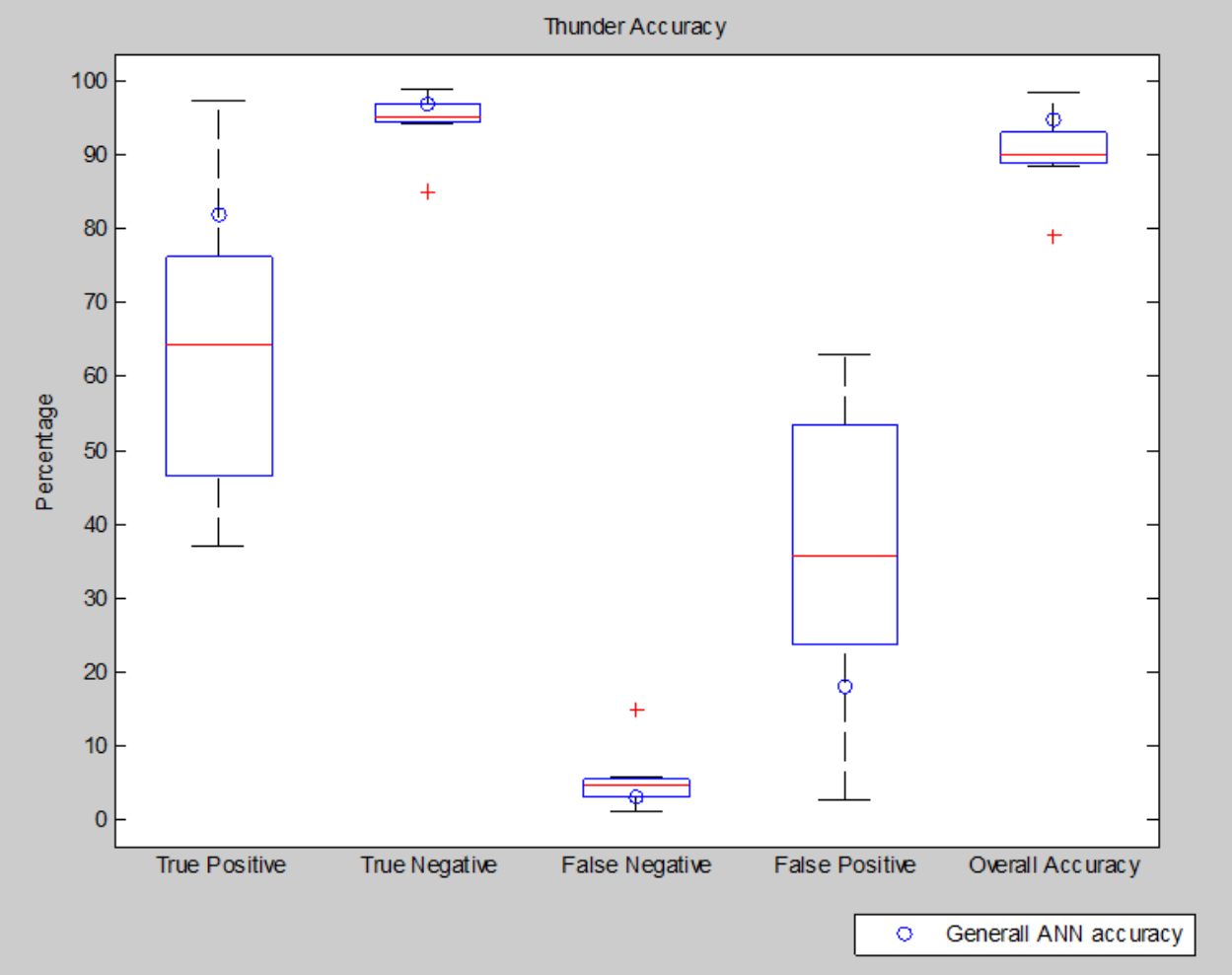

Figure 39: Human to ANN Accuracy Comparison: Thunder

7.12 FINAL ANN

Figure 40 shows the confusion matrix of an ANN trained using the most effective training techniques discussed in this study. Including the 21 additional metrics examined in this study increases ANN accuracy by 28.4% across all data libraries as shown in Table 10. Exclusion of mixed blasts did not hinder blast classification in less windy situations; however, it did hinder the ANN's ability to properly classify mixed blasts. The "Max" method of ANN evaluation was preferred to the "Round" method because it produces no null classifications and exhibited higher overall accuracy. Excluding wind from training did not impact blast classification under less windy circumstances; however, when wind bypassed the measures meant to exclude it, there was a high probability that it will lead to a blast false positive without wind training. The ANN was found to be highly accurate when the ratio of blast to wind was over 10 dB; however, low ratios of blast and wind will cause a large spike in blast misclassifications. And lastly, the new ANN was found to have sound classification accuracy comparable to the test subjects under identical circumstances. Table 40 shows a 28.4% increase in overall accuracy and a 10.6% increase in blast true positive accuracy from the old UPITT classifier to the new one.

blast	wind	mach	air craft	vehicle	thunder	
14159	131	127	3	40	411	4.79%
150	12440	14	28	270	39	3.87%
120	8	1752	17	81	21	12.36%
4	9	21	3135	337	45	11.72%
46	278	63	374	9241	188	9.31%
232	84	23	44	156	3306	14.02%
3.75%	3.94%	12.40%	12.94%	8.73%	17.56%	92.90%

Blast

14159	712	96.25%
552	31974	97.82%
2.18%	3.75%	97.33%

Air Craft

3135	416	87.06%
466	43380	99.05%
0.95%	12.94%	98.14%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Wind

12440	501	96.06%
510	33946	98.55%
1.45%	3.94%	97.87%

Vehicle

9241	949	91.27%
884	36323	97.45%
2.55%	8.73%	96.13%

Machine Gun

1752	247	87.60%
248	45150	99.46%
0.54%	12.40%	98.96%

Thunder

3306	539	82.44%
704	42848	98.76%
1.24%	17.56%	97.38%

Figure 40: ALL Metric ANN Confusion Matrix (Refined Classifications)

Table 10: ANN Performance Comparison.

	All New Metrics		Old Metrics	
	Blast TP %	Overall Accuracy %	Blast TP %	Overall Accuracy %
SERDP	98.7%	98.7%	98.0%	98.0%
Base 1	96.7%	92.9%	81.0%	63.0%
Base 2	96.0%	90.5%	84.0%	70.5%
All Combined	97.3%	92.9%	86.7%	64.50%

8.0 ANN STRUCTURE ECONOMIZATION

Although the final ANN provided very high blast and and high overall accuracy, its usage may prove cumbersome to less powerful computer systems. It is possible that not every metric may be necessary and the ANN's structure may be an unnecessarily complex (ie. it may have more neurons or layers than needed). To determine the if the number of metrics, hidden layers, and neurons could be downsized without compromising accuracy, two ANN structure economization methods were utilized: the Forward Sequential Selection and Network Pruning[[23] & [24]].

8.1 FORWARD SEQUENTIAL SELECTION (FSS)

Although many new metrics were considered, it was expected that some were more useful than others. For that reason, a Forward Sequential Selection (FSS) was performed to determine the order of metrics from most to least important and how they impacted the overall accuracy of the classifier. From the list generated by the FSS, a reduced set of metrics could be extracted which would allow the classifier to provide high accuracy at reduced computational expense.

First, "N" ANN's were trained using each of the metrics individually. Once every ANN had been evaluated, the metric that produced the best overall test accuracy was saved into the "include metric" vector.

Next, "$N-1$" ANN's were made using two metrics: the best one from the previous round and each remaining metric individually. The best combination of metrics was then saved in the "include metric" vector.

Next, "$N-2$" ANN's were made using three metrics the best combination of two from the previous round and each remaining metric individually. This process continued until the one final ANN was made using all of the metrics. Once completed, the "include metric" vector contained all of the metrics in order of most influential to least.

By examining the increase in overall test accuracy in Figures 41 and 42, it was possible to track the increase in performance as a function of metrics added. When the overall test accuracy failed to increase, the metrics used up until that point were assumed to be the most necessary.

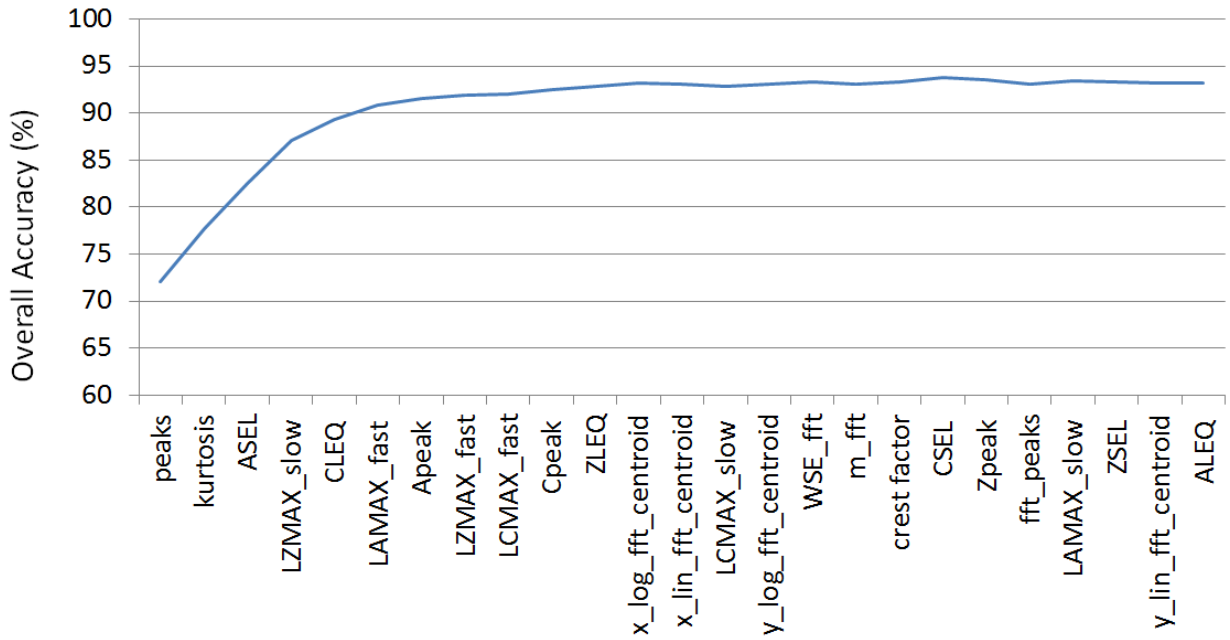

Figure 41: Base 1 data set forward sequential selection.

Figures 41 and 42 display how the overall accuracy fails to increase after the eight most important metrics were added. The SERDP library was not included as part of the FSS since it didn't have recordings for all six classes of noise. Unfortunately, the first two data sets had eight different most important metrics. Although blast noise types were similar for both sites, the presence (or lack there of) of other noise sources such as different

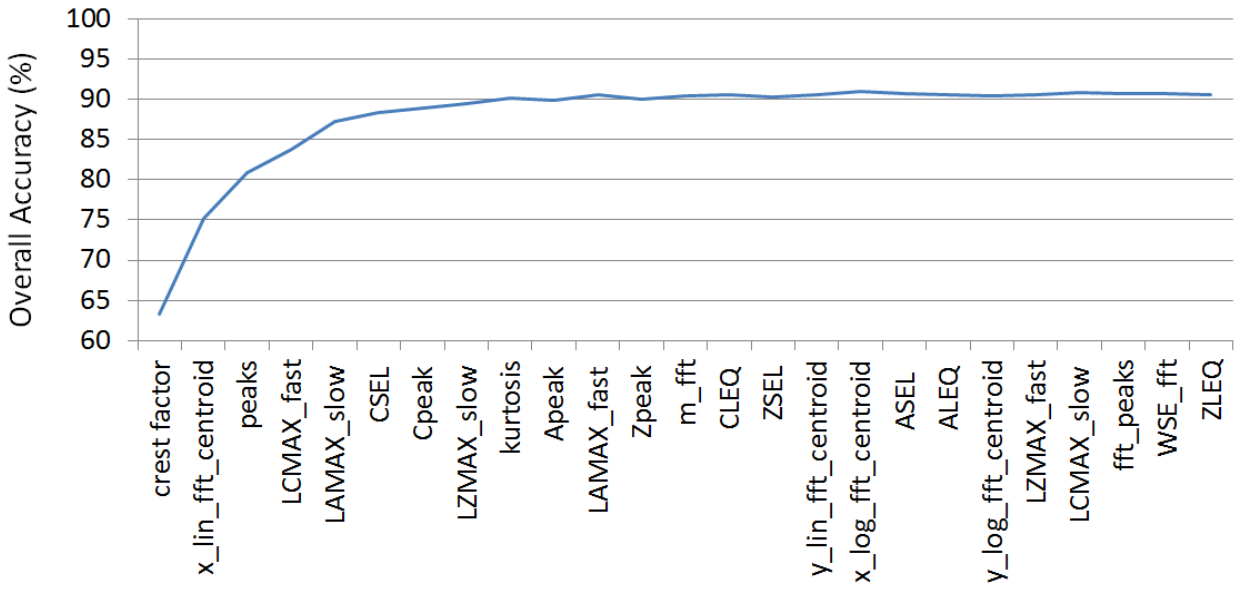

Figure 42: Base 2 data set forward sequential selection.

types of vehicles (tanks, cars, atv, etc.) and aircraft (helicopter, distant jets, and airplanes) were possible reasons for the different sets of most important metrics. This implies that efficient ANN performance could be reliably obtained with site specific training; however, such training methods would severely limit widespread adaptation of the product.

To make the classifier more universal, an FSS was performed on a mixture of all the data to obtain a general ANN to be used at all sites. Table 11 shows that the "FSS ANN," once trained, had an overall accuracy of 90.37% with the inclusion of only the best eight metrics. Figure 43 displays the confusion matrix of the "Eight Metric ANN" that uses the eight metrics from the FSS while Table 12 compares it's accuracy to that of the original UPITT classifier and the ALL metric ANN.

Table 11: Eight FSS Metric ANN performance.

Metrics	ASEL	Kurtosis	X linear Centroid	ZSEL	LA Max Fast	A peak	Spectral Slope	Crest Factor
Overall Accuracy	61.90%	76.30%	83.30%	86.70%	88.80%	89.79%	89.80%	90.37%

blast	wind	mach	air craft	vehicle	thunder	
13911	203	174	3	65	557	6.72%
274	12303	11	7	204	11	3.96%
137	5	1503	22	94	59	17.42%
7	12	49	2971	434	28	15.14%
64	335	93	571	9018	393	13.90%
318	92	170	27	310	2962	23.64%
5.44%	5.00%	24.85%	17.50%	10.93%	26.13%	90.02%

Blast

13911	1002	94.56%
800	31684	96.93%
3.07%	5.44%	96.20%

Air Craft

2971	530	82.50%
630	43266	98.79%
1.21%	17.50%	97.55%

Rubric

#TP	#FP	%TP
#FN	#TN	%TN
%FP	%FN	%OV AC

Wind

12303	507	95.00%
647	33940	98.53%
1.47%	5.00%	97.57%

Vehicle

9018	1456	89.07%
1107	35816	96.09%
3.91%	10.93%	94.59%

Machine Gun

1503	317	75.15%
497	45080	99.30%
0.70%	24.85%	98.28%

Thunder

2962	917	73.87%
1048	42470	97.89%
2.11%	26.13%	95.85%

Metrics

- ASEL
- Kurtosis
- X lin centroid
- ZSEL
- LAMAX fast
- A peak
- Spectral Slope
- Crest factor

Figure 43: Eight FSS Metric ANN Confusion Matrix

Table 12: All/Eight FSS/Old Metric ANN performance.

	ALL Metrics		Eight FSS Metrics		Old Metrics	
	Blast TP %	Overall Accuracy %	Blast TP %	Overall Accuracy %	Blast TP %	Overall Accuracy %
All Combined	97.3%	92.9%	95.28%	90.37%	86.7%	64.50%

8.1.1 Wind Speed Analysis

After the completion of the SNR, it was noted that blast misclassification was a function of the ratio of blast to wind but a question arose as whether it was also a function of wind speed, since wind noise is proportional to wind speed. And if it was, would the new classifier be able to improve upon it. Figures 44 and 45 display samples of files collected between the hours of 9 am to 6 pm on August 21st and 8 am to 6 pm on October 30th 2013, respectively.

Tables 13 and 14 show the number of correct and incorrect classifications for each day respectively. For both the old and new classifier, the ratio of correct classifications to incorrect decreased as wind speed increased. The new classifier, however, had larger ratios of correct to incorrect classification for all but one wind speed range where the ANN's differed by one classification.

In summary, a positive correlation between wind speed and decreased correct to incorrect classification ratios. For both days, the new classifier increased overall accuracy above 92% and in many cases almost doubled the ratio of correct to incorrect classifications.

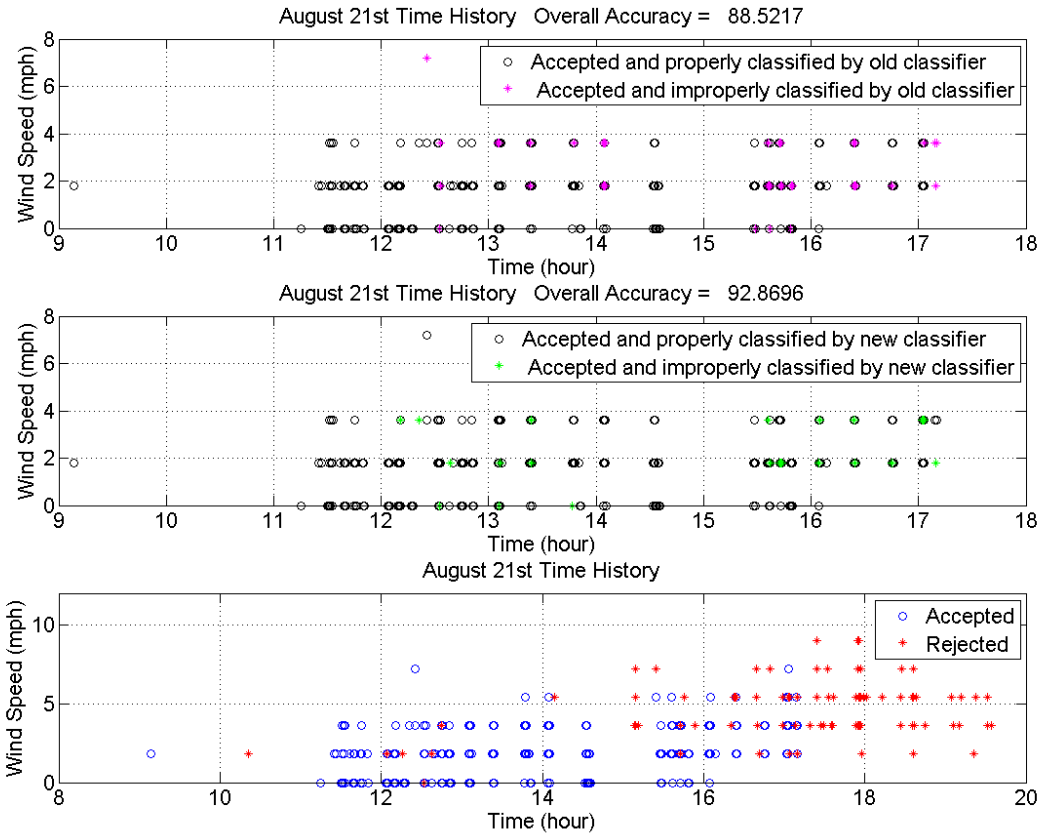

Figure 44: Classification of Blast Files Under Various Wind Speeds For August 21st (Eight FSS Metrics)

Table 13: Classifications of Blast Files Under Various Wind Speeds For Aug 21st (Eight FSS Metrics).

Wind Speed (mph)	# Correct (Old)	# Incorrect (Old)	Total	# Correct (New)	# Incorrect (New)	Total
0	149	6	155	152	3	155
1.125	246	30	276	260	16	276
2.25	0	0	0	0	0	0
3.375	101	23	124	109	15	124
4.5	13	5	18	12	6	18
5.625	0	0	0	0	0	0
6.75	0	2	2	1	1	2
7.875	0	0	0	0	0	0
9	0	0	0	0	0	0
10.125	0	0	0	0	0	0
11.25	0	0	0	0	0	0

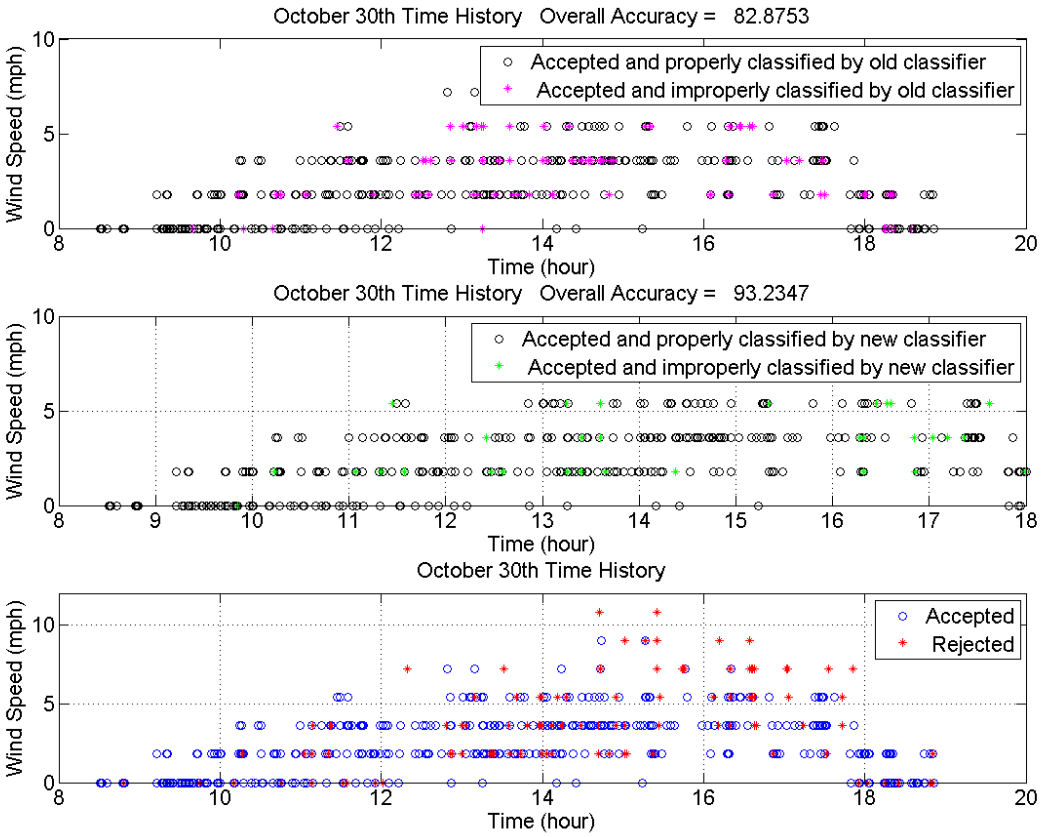

Figure 45: Classification of Blast Files Under Various Wind Speeds For October 30th (Eight FSS Metrics)

Table 14: Classifications of Blast Files Under Various Wind Speeds For Oct 30th (Eight FSS Metrics).

Wind Speed (mph)	# Correct (Old)	# Incorrect (Old)	Total	# Correct (New)	# Incorrect (New)	Total
0	124	10	134	133	1	134
1.125	127	27	154	141	13	154
2.25	0	0	0	0	0	0
3.375	102	24	126	117	9	126
4.5	32	19	51	43	8	51
5.625	0	0	0	0	0	0
6.75	6	0	6	5	1	6
7.875	0	0	0	0	0	0
9	1	1	2	2	0	2
10.125	0	0	0	0	0	0
11.25	0	0	0	0	0	0

8.1.2 SNR for Eight FSS Metric ANN

The SNR analysis was revisited for the Eight FSS Metric ANN to test it's sensitivity to wind. According to Figure 46, at a desired SNR of 0 dB it showed a blast accuracy of 73.2%. Similarly to the full metric ANN, at a desired SNR of 10 dB the Eight FSS Metric ANN had a blast accuracy of 97.5% and a blast accuracy of 99.9% at a desired SNR of 40 dB. Although the Eight FSS Metric ANN was as insensitive to wind at high signal to noise ratios, it proved to be more sensitive at lower signal to noise ratios. This means that under severely windy conditions the Eight Metric ANN would perform worse than the full ANN, but still better than the Chan classifier by 11.00% and the Lopatka classifier by 23.65% at 10 dB.

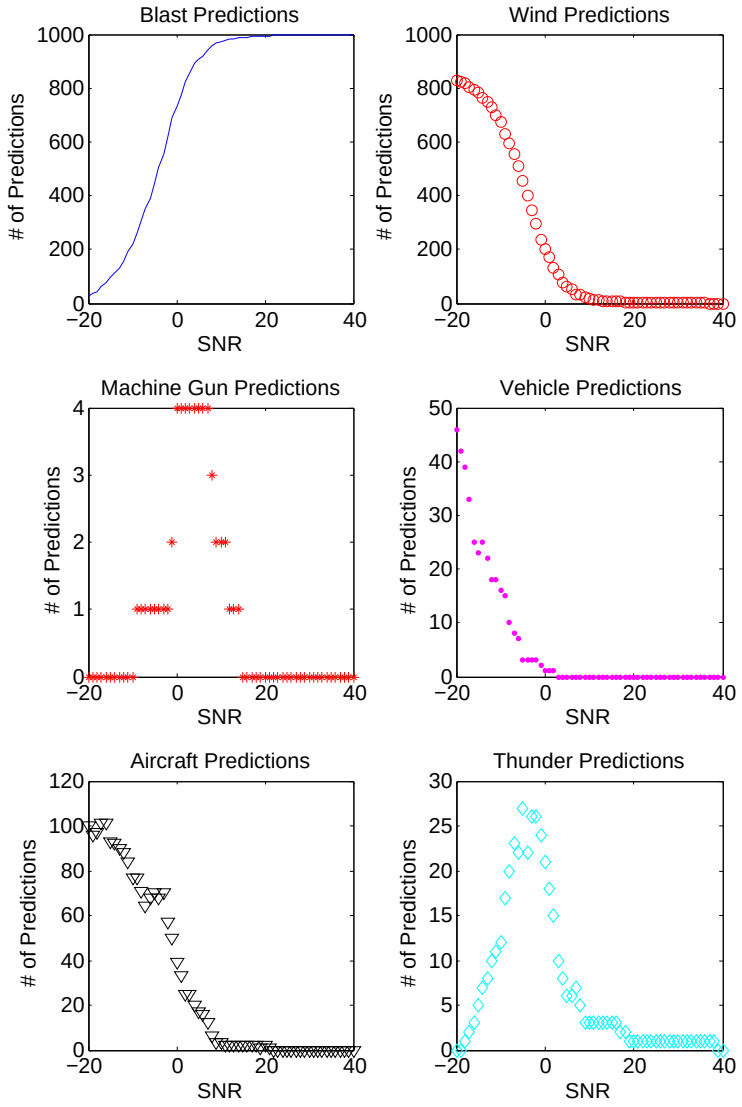

Figure 46: Classifier Performance at Various Signal (pure blast) to Noise (pure wind) Ratios for the Eight FSS Metric ANN

8.1.3 Human Classification Comparison (FSS)

Figure 47 shows that the both the All Metric and Eight FSS Metric ANNs were more accurate than most test subjects at correctly classifying blast true positives and false positives; however, the All Metric ANN performed better than the Eight FSS Metric ANN on both accounts. The Eight FSS Metric ANN was better at discerning blast true negatives and false negatives and ultimately performed just as well overall as the All Metric ANN and the median test subject for the same 1,000 random samples used in the creation of Figure 47; making the Eight FSS metric ANN very much on par with the All metric ANN and the median test subject.

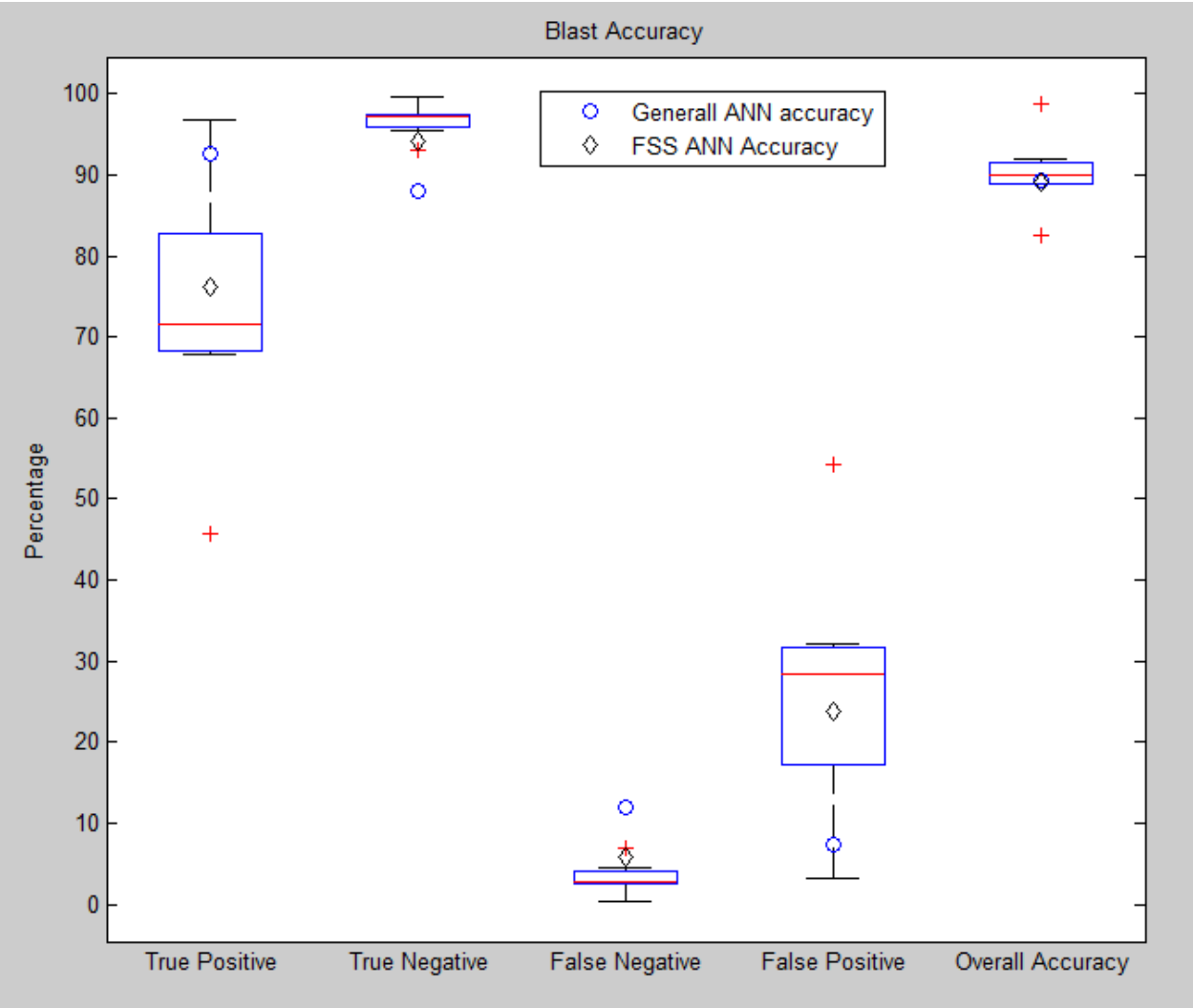

Figure 47: Human to FSS ANN Accuracy Comparison: Blast

The Eight FSS Metric ANN overall performed worse than the median test subject and the All Metric ANN at wind, machine gun, and vehicle classification; however its accuracy for each class was still at roughly 90%. It performed as well as the median test subject for aircraft and thunder classification but slightly less effectively as the ALL metric ANN for both cases.

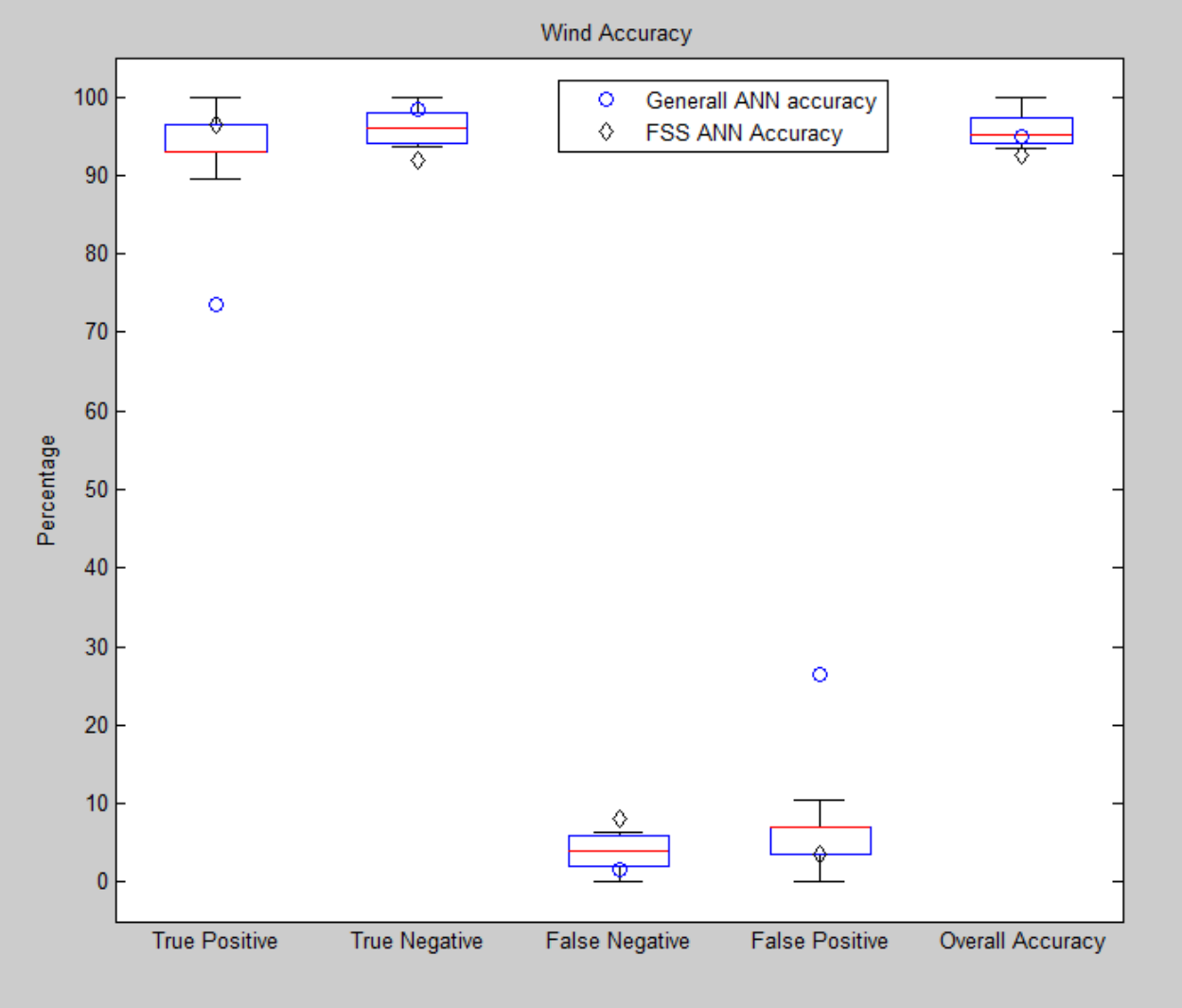

Figure 48: Human to FSS ANN Accuracy Comparison: Wind

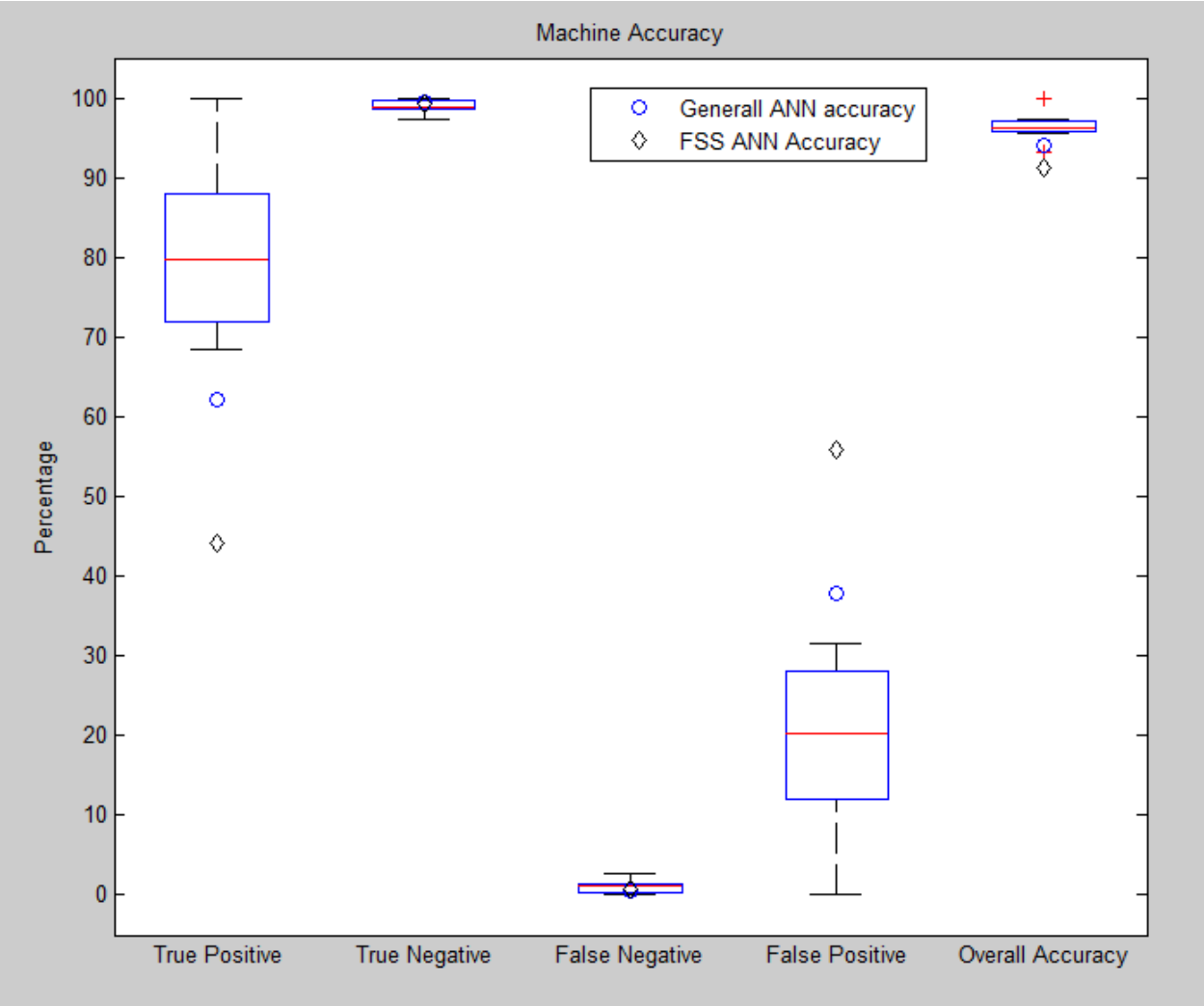

Figure 49: Human to FSS ANN Accuracy Comparison: Machine Gun

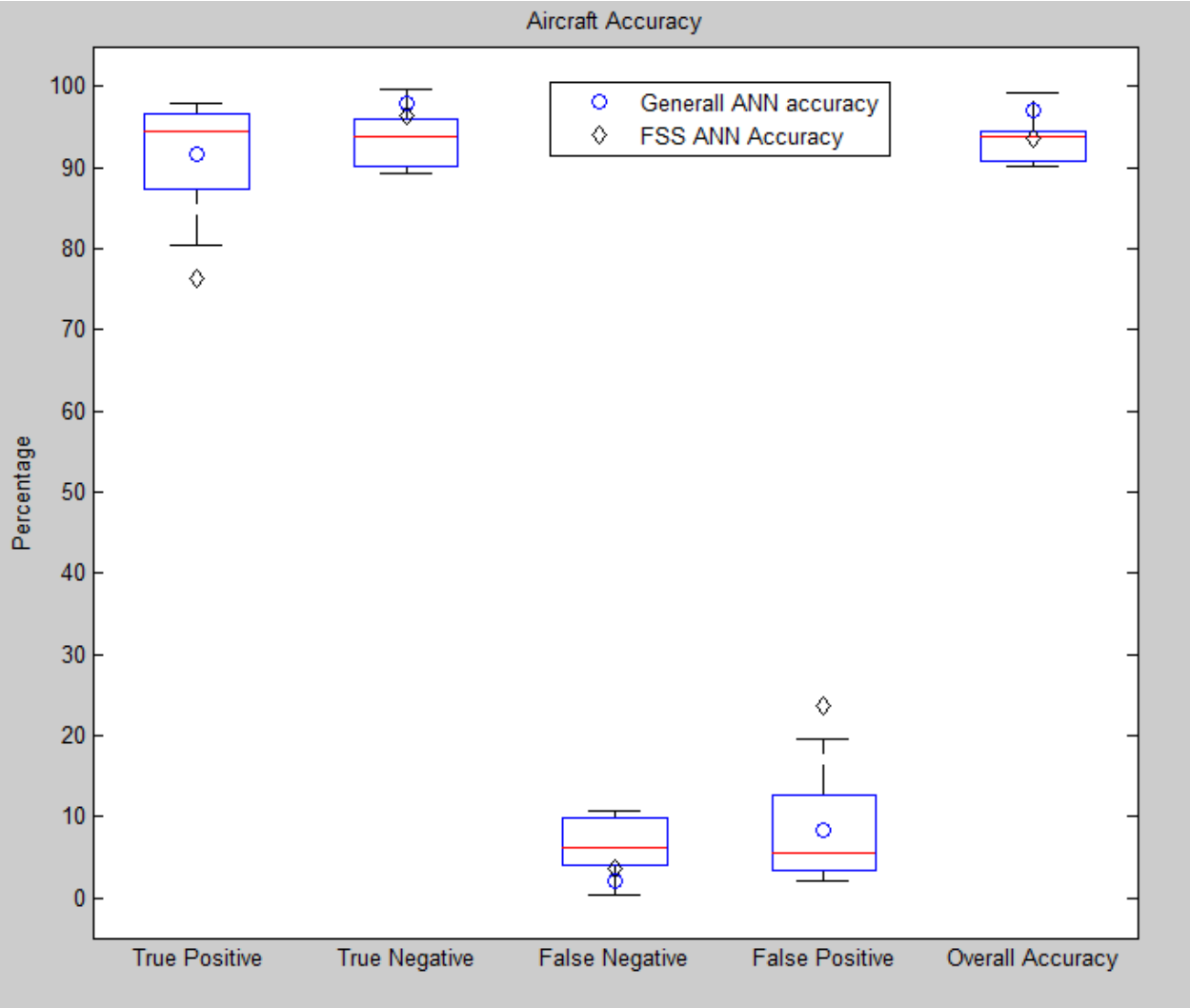

Figure 50: Human to FSS ANN Accuracy Comparison: Air Craft

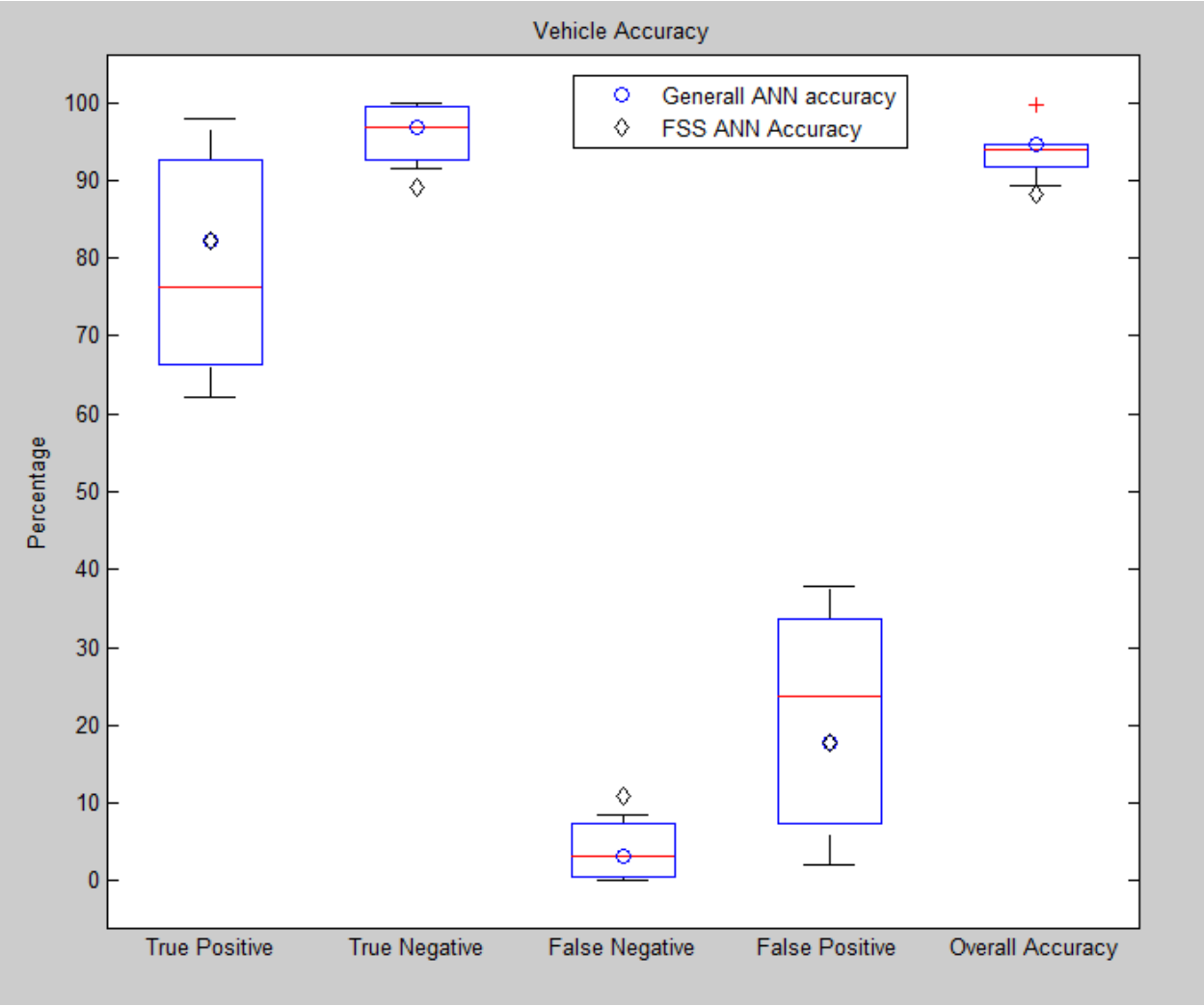

Figure 51: Human to FSS ANN Accuracy Comparison: Vehicle

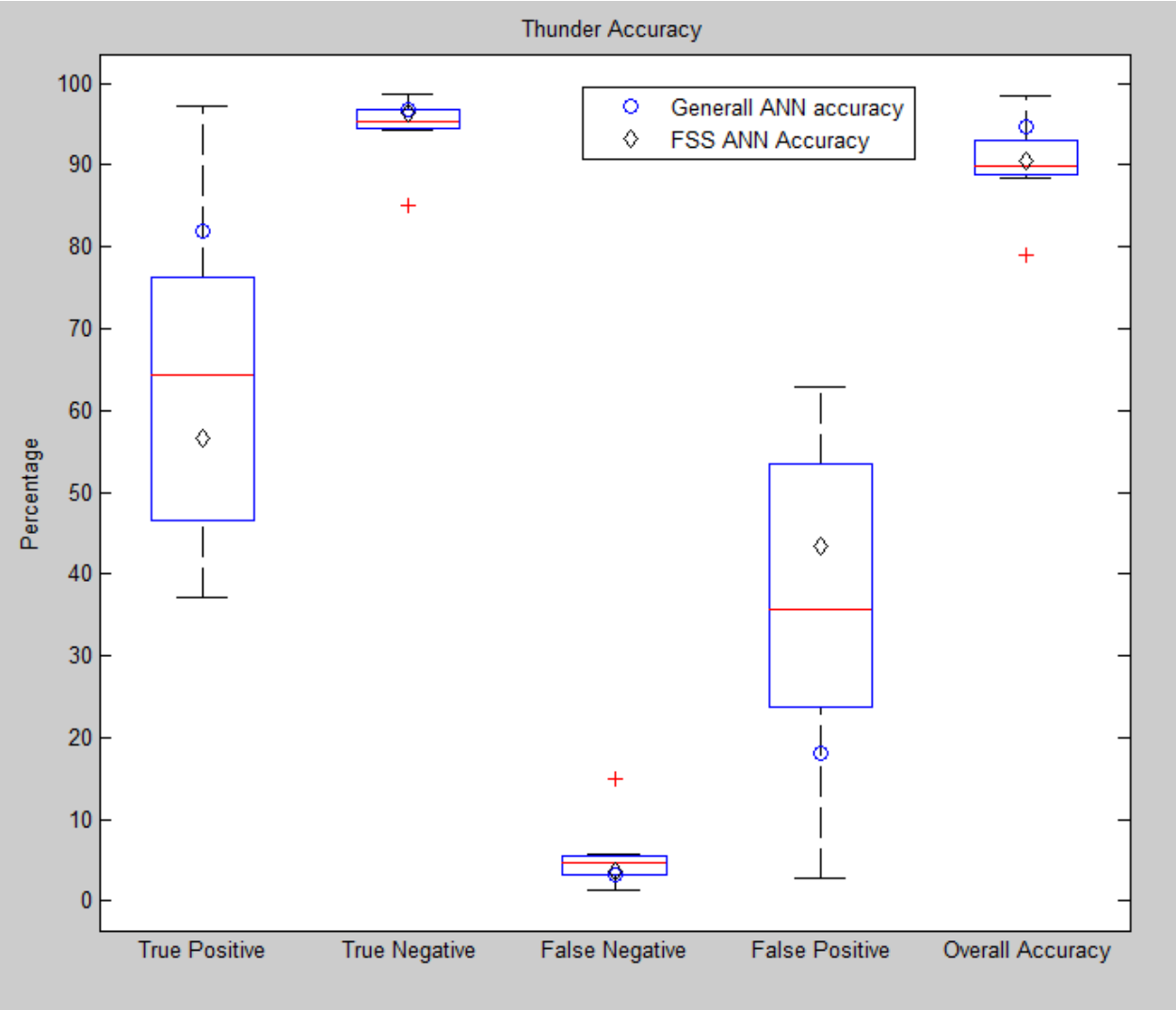

Figure 52: Human to FSS ANN Accuracy Comparison: Thunder

Ultimately the FSS Metric ANN performed as well as the median test subject and the ALL Metric ANN for blast classification, which was the focus of the study. By leaving out the other seventeen metrics, it was possible to obtain 90% overall accuracy and retain high blast, aircraft, and thunder accuracy at the cost of the other three sound classes.

8.1.4 Network Pruning

Using the FSS metrics, five ANN's were were trained using between one (1) and ten (10) layers with between one(1) and ten (10) neurons each totaling 500 ANN's. Figures 53 and 54 show that even ANN's with large numbers of neurons and layers do not exceed 91% overall classification accuracy. There were two instances where ANN's with 2 layers achieved 100% accuracy. To discover whether these accuracies were statistical anomalies, the mean overall accuracies and their standard deviations were calculated.

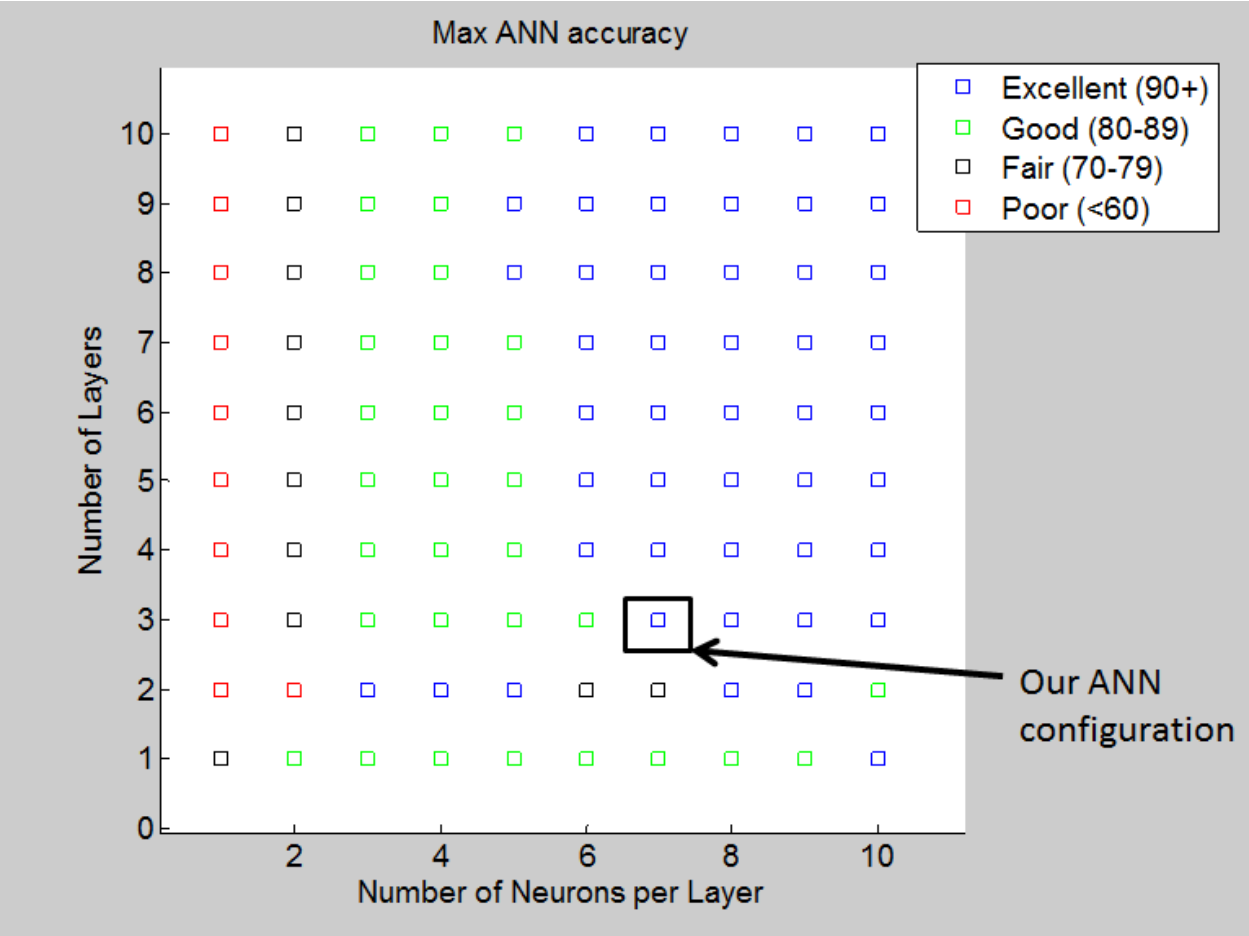

Figure 53: Max ANN Accuracy

Number of Layers										
10	32	76	80	86	88	90	90	90	91	90
9	31	75	80	86	90	90	90	90	91	91
8	55	76	80	85	90	90	90	91	91	91
7	56	76	80	86	89	90	90	90	91	91
6	56	75	81	87	89	90	90	91	91	91
5	56	76	81	85	89	90	91	91	91	91
4	56	76	81	85	89	90	90	90	91	91
3	56	76	81	85	89	89	90	90	90	90
2	54	49	100	92	99	79	78	100	95	84
1	70	80	83	85	87	88	87	88	89	90
	1	2	3	4	5	6	7	8	9	10

Number of Neurons per Layer

Figure 54: Max ANN Accuracy Values

The blue colored region in Figure 55 shows that only complex ANN's achieve a mean accuracy greater that 90%. Each of the ANN's with two layers and 100% accuracy had either poor or fair average accuracy and accuracy standard deviations greater than 15%.

During the study, a phenomenon was discovered. Sometimes for less complex ANN configurations, the ANN classification result for every class was equal to 0.5. Both the max and round method interpreted those outputs as 6×1 vectors populated exclusively by one's. The overall accuracy was computed as the number of true positives divided by the total number of incidents (since in our data set, every file was a true positive of one of the six noise types). ANN's which produce this output did so for every input, this yielding an incorrect overall accuracy of 100%.

ANN's in the blue region of Figure 57 had overall accuracy standard deviations less than 1%. High mean overall accuracies coupled with low overall accuracy standard deviations imply that ANN's trained with those hidden layer and neuron configurations will consistently yield ANN's with the highest possible accuracy.

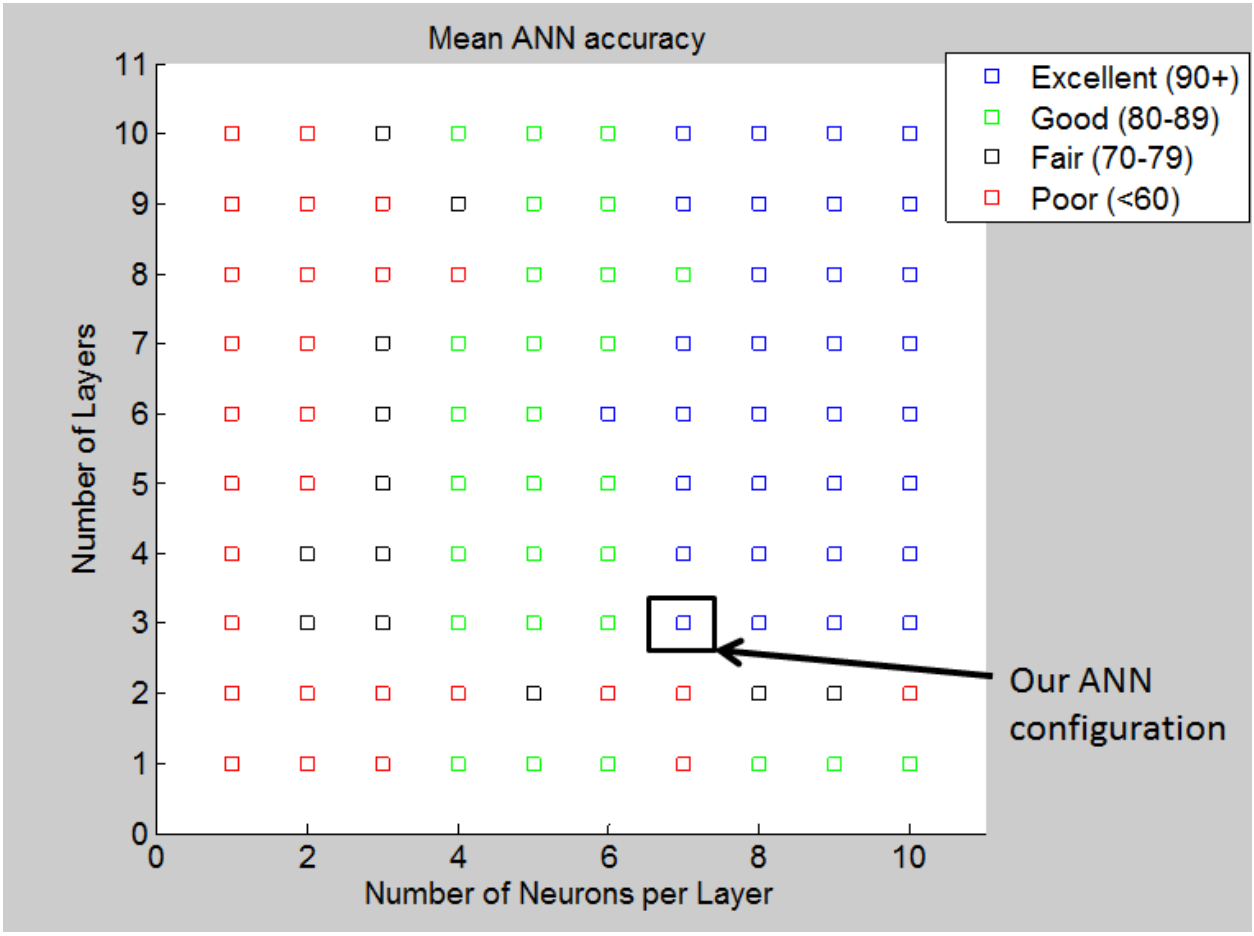

Figure 55: Mean ANN Accuracy

Number of Layers										
10	31	40	77	85	85	89	90	90	90	90
9	31	40	69	73	89	89	90	90	90	90
8	36	54	60	63	88	88	89	90	90	90
7	45	49	74	86	89	89	90	90	90	91
6	53	49	75	86	89	90	90	90	90	90
5	45	49	75	85	89	89	90	90	90	90
4	56	76	75	80	88	89	90	90	91	90
3	41	74	79	82	88	89	90	90	90	90
2	37	28	33	45	79	27	45	72	76	68
1	41	66	51	80	83	87	61	81	82	88
	1	2	3	4	5	6	7	8	9	10

Number of Neurons per Layer

Figure 56: Mean ANN Accuracy Values

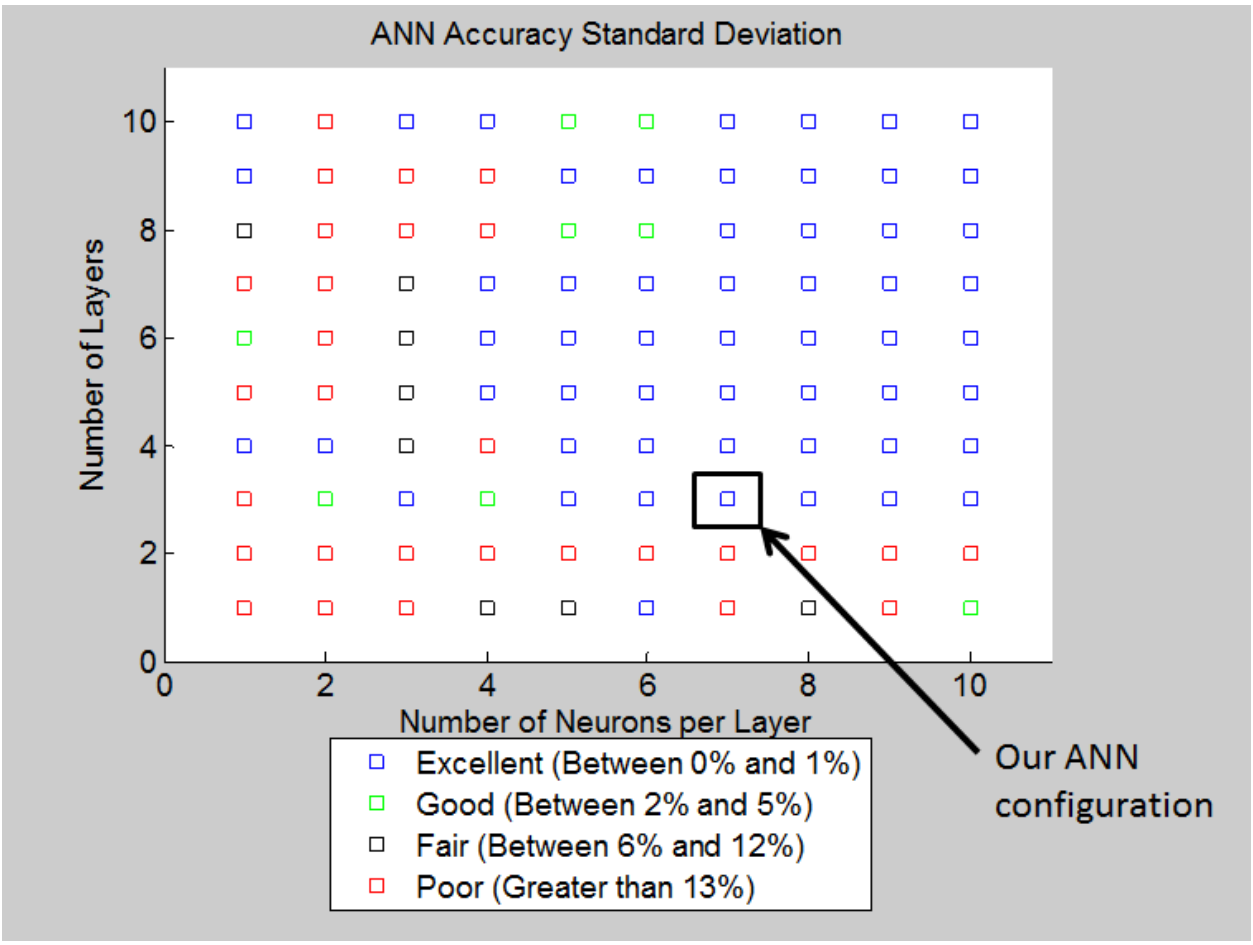

Figure 57: STD ANN Accuracy

Number of Layers										
10	0.50	20.04	1.45	1.42	4.26	2.26	0.46	0.33	0.49	0.21
9	0.20	19.95	21.19	23.58	0.40	0.20	0.38	0.33	0.24	0.37
8	10.88	22.42	26.36	28.95	1.90	3.85	0.39	0.10	0.34	0.47
7	12.91	24.42	10.50	0.29	0.56	1.22	0.42	0.47	0.51	0.23
6	2.58	24.12	10.64	0.68	0.40	0.24	0.21	0.37	0.35	0.34
5	12.91	24.45	10.75	0.92	0.33	0.56	0.57	0.31	0.30	0.37
4	0.34	0.18	10.43	12.70	0.26	0.40	0.20	0.11	0.43	0.17
3	13.60	3.65	1.18	4.00	0.54	0.36	0.36	0.44	0.14	0.28
2	19.27	15.28	38.93	31.51	27.73	29.61	23.71	31.88	25.62	17.28
1	24.05	19.71	28.87	8.83	6.71	0.46	33.06	12.34	14.67	2.52
	1	2	3	4	5	6	7	8	9	10

Number of Neurons per Layer

Figure 58: STD ANN Accuracy Values

9.0 ANN OUTPUT DISTRIBUTION

Once the economized ANN was found through the FSS and Network Pruning validated the size of the ANN, the output distribution was analyzed to find patterns between positive, negative, correct and incorrect classifications. The first goal of this analysis was to find a clear distinction between positive and negative classifications.

"Round method" assumed that the ANN's positive prediction (ie. what it thinks caused the sound) had an output greater than 0.5 and that there was exactly one sound source that met this criteria; however, the appearance of null classifications showed that some ANN classifications had outputs which were all less than 0.5 which resulted in increased false negative classifications. "Max Method" assumed that the positive classification was the highest of the six outputs; however, this method forced the ANN to make a decision for every sound file. By taking away the ability to make null classifications; the ANN was forced to classify unknown sound sources which could result in increased false positive classifications.

Firstly, patterns between positive and negative classifications were analyzed using boxplots of the ANN outputs as well as the receiver operator curves. It could then be determined if implementing a different threshold (ie. different from the round method's 0.5 threshold) could increase ANN classification accuracy while limiting the number of false negatives, null classifications, and double positives (ie. cases where two different outputs both cross the threshold). Secondly, patterns between the highest and the second highest ANN output were analyzed to see if the later could be thought of as a "second best guess". Should a pattern be found when the second guess was right, logic could be implemented to correctly classify files that would otherwise be misclassified.

Ideally, the logsig transfer function in the third hidden layer should squash all ANN outputs to values between 0 and 1; however when the raw outputs of the ANN were more closely examined some outputs were found to be outside of those bounds. This wasn't noticed before because both the "round" and "max" methods rounded or forced the outputs to be either to 0 or 1. Before metrics were input into an ANN, they were normalized relative to the largest values of that metric experienced during ANN training. When new values of those metrics were input into the system that exceed the maximum values seen in training, the ANN output can be greater than 1. Similarly, if new values of those metrics were input into the system that are smaller than the minimum values seen in training, the result could be less than 0. To address these issues, a saturating linear transfer function was applied to the ANN output to constrict them between 0 and 1.

Figure 59 shows boxplots of the file ANN output in terms of blast classifications. The x-axis refers to the number of the class(ie. blast = 1, wind = 2, machine gun = 3, aircraft = 4, vehicle = 5, and thunder = 6). Blast true positives have ANN output are tightly grouped between 1 and 0.9 with outliers stretching down to 0.3; indicating high confidence in the blast true positive classifications. False positives were not as tightly centered and had a much wider range of ANN outputs which stretched from 1 and 0.3. Blast false negatives were mostly attributed to wind; indicated by the wide array of ANN outputs shown in the bottom left boxplot of Figure 59. This was thought to be a byproduct of training with mixed blasts. The ANN was trained to consider any file with any trace of blast a blast file, regardless of how noisy it was, in order to maximize the number of properly classify mixed blasts. Doing so increased the ANN's ability to accurately classify mixed blasts dramatically. A unfortunate side effect was that some files with very small blast to wind ratios were human classified as blasts for the sake of training but were too noisy for the classifier accurately identify.

Figure 60 shows a receiver operating characteristic (ROC) curve for the binary classification of blast noise as a the threshold of the classifier changed. A particular threshold can be chose to achieve the desired balance between true positive and false positive rates. At a threshold of 0.5 (typical), the true positive accuracy is 85.06% and the false positive rate is 2.90%. At a threshold of 0.9, where Figure 59 shows that the majority

of the blast true positive output lie, Figure 60 displays a true positive accuracy of 73.09% with a false positive rate of 0.99%. Lastly at a threshold of 0.3, the bottom most ANN output for blast true positives, there was a true positive accuracy of 94.84, but a false positive rate of 14.88%. If only the blast output were to be utilized, a threshold of 0.5 seems to provide a reasonable ratio of true positive to false positive rates.

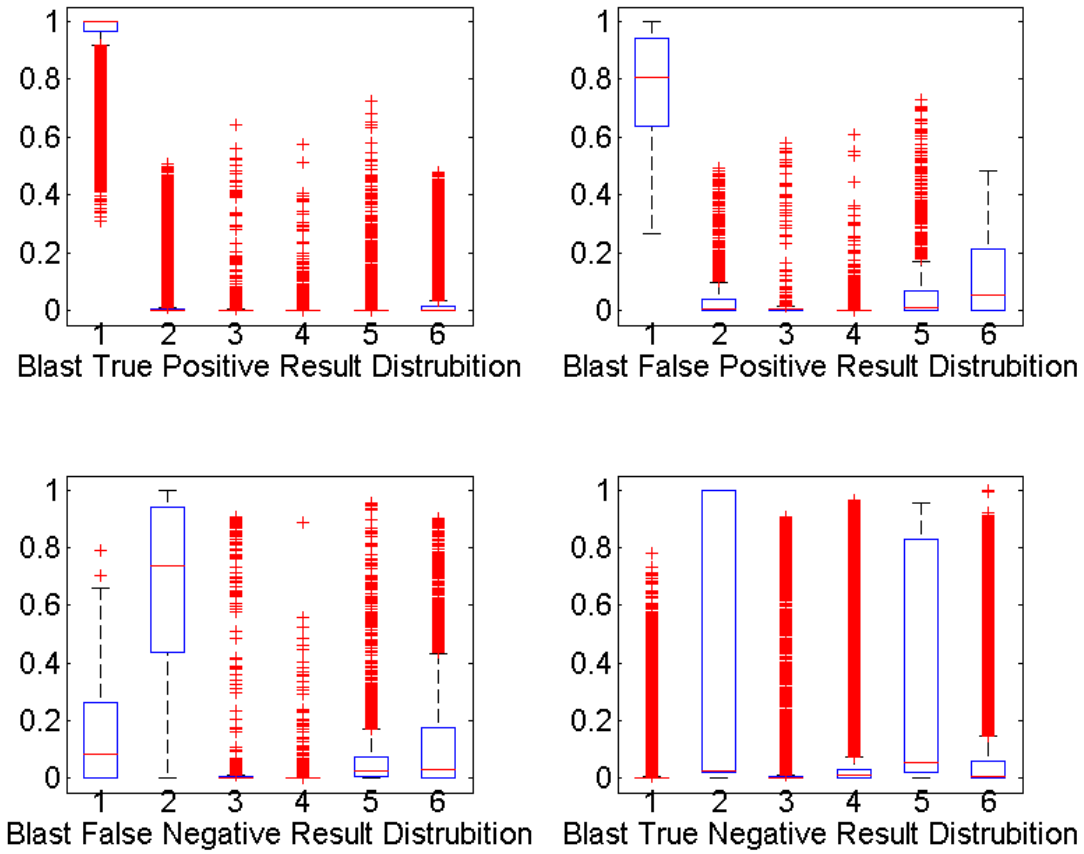

Figure 59: Blast Result Distribution

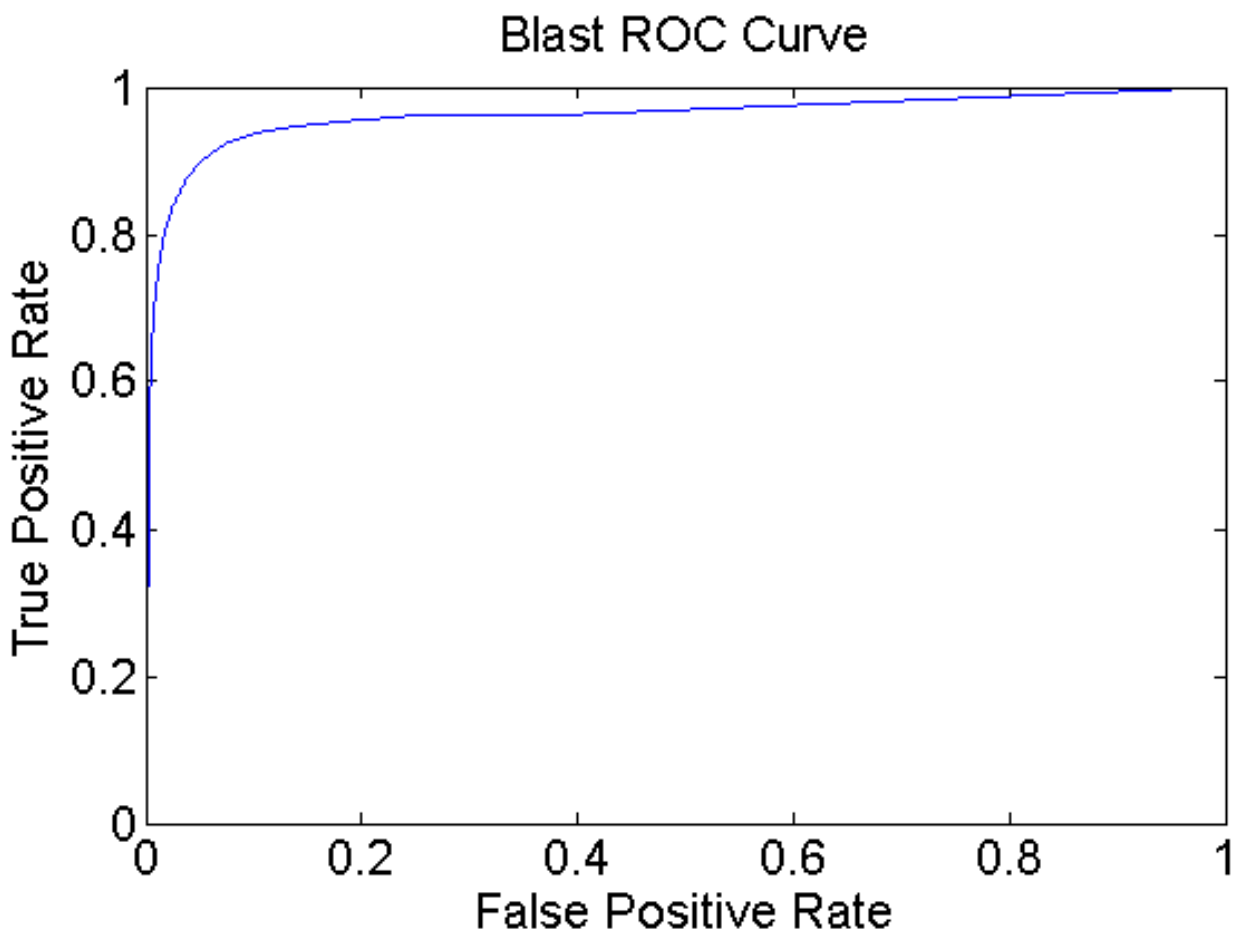

Figure 60: Blast ROC Curve

Figures 61 shows a much tighter grouping of wind true positives with a median ANN output of 1 with outliers reaching down to 0.35. False positive ANN outputs stretched from 1 to 0.27. Figure 62 shows that at a threshold of 1, the true positive rate was 77.78% and a false positive rate of 0.48%. At a threshold of 0.35, the lowest wind true positive output, there was a true positive rate of 96.98% and false positive rate of 7.77%. Lastly, at a threshold of 0.5, the lowest wind true positive output, there was a true positive rate of 96.56% and false positive rate of 6.25%.

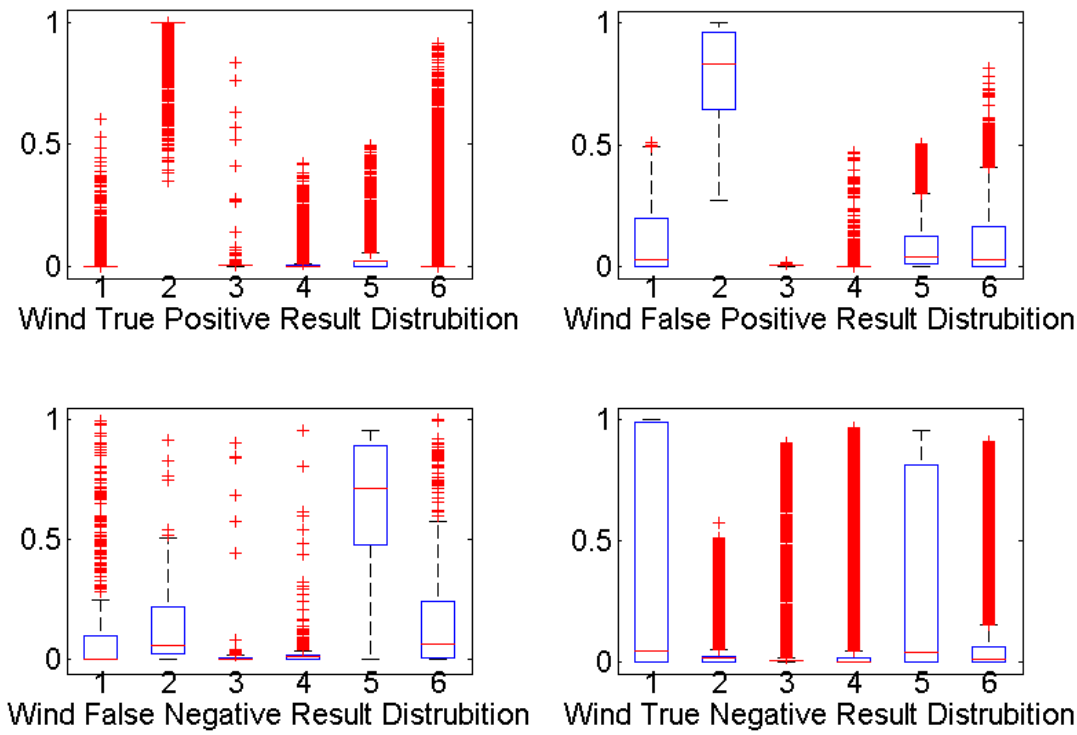

Figure 61: Wind Result Distribution

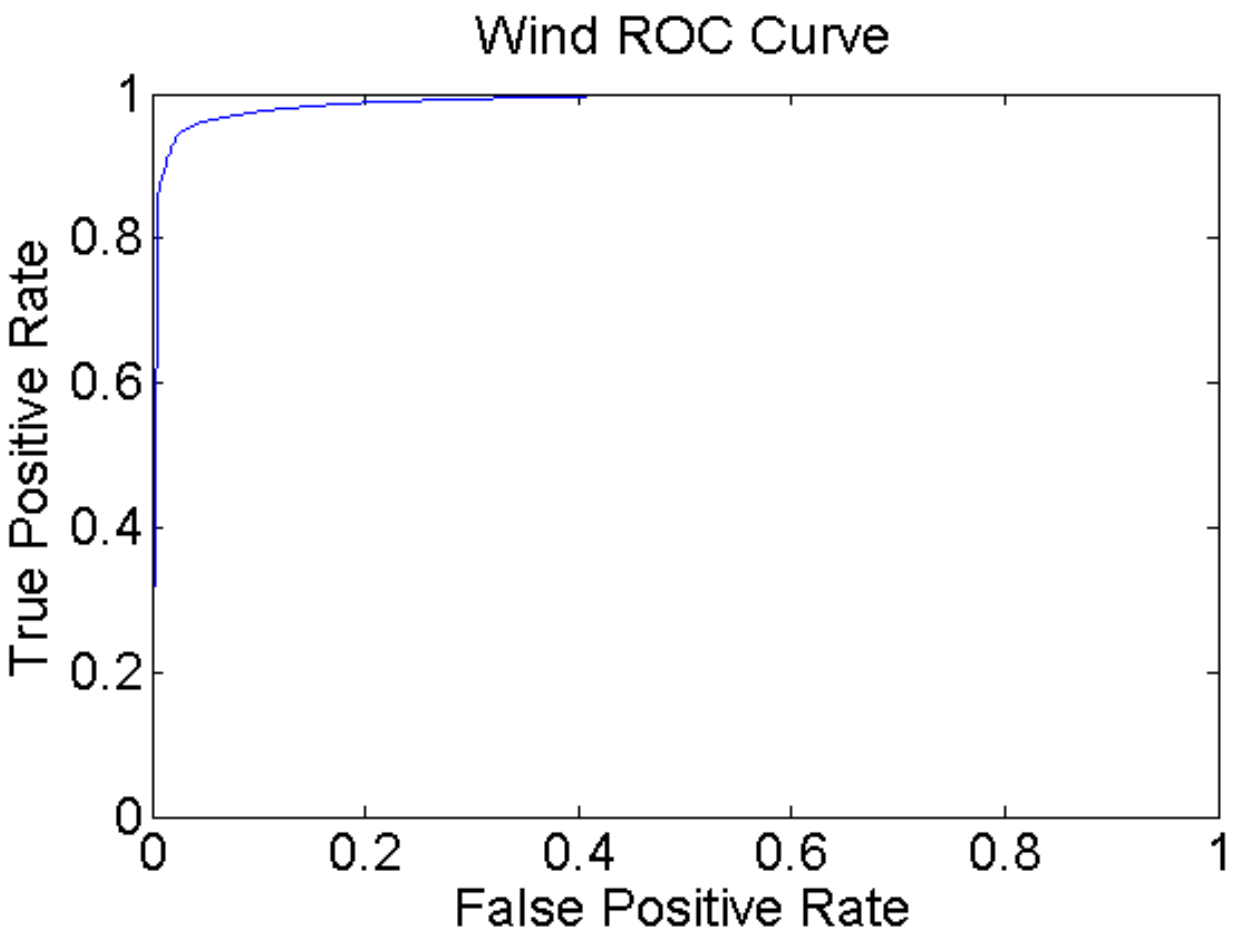

Figure 62: Wind ROC Curve

Unlike blast and wind, machine gun ANN outputs were less tightly centered at the median. Figure 64 shows that at a threshold 0.5, there was a true positive rate of 46.45% and a false positive rate of 0.45%. At a threshold of 0.9, there was a true positive rate of 21.1% and a false positive rate of 0.15%. This indicates that machine gun files are particularly difficult to properly classify using a threshold method.

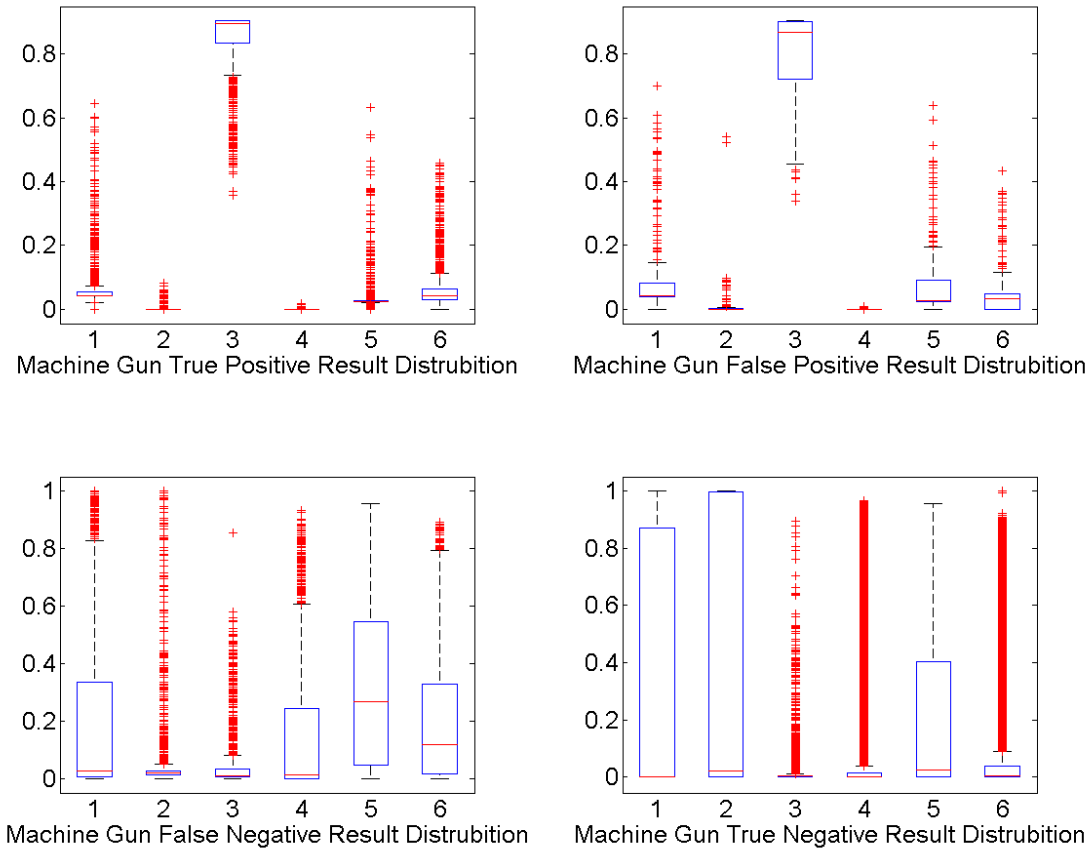

Figure 63: Machine Gun Result Distribution

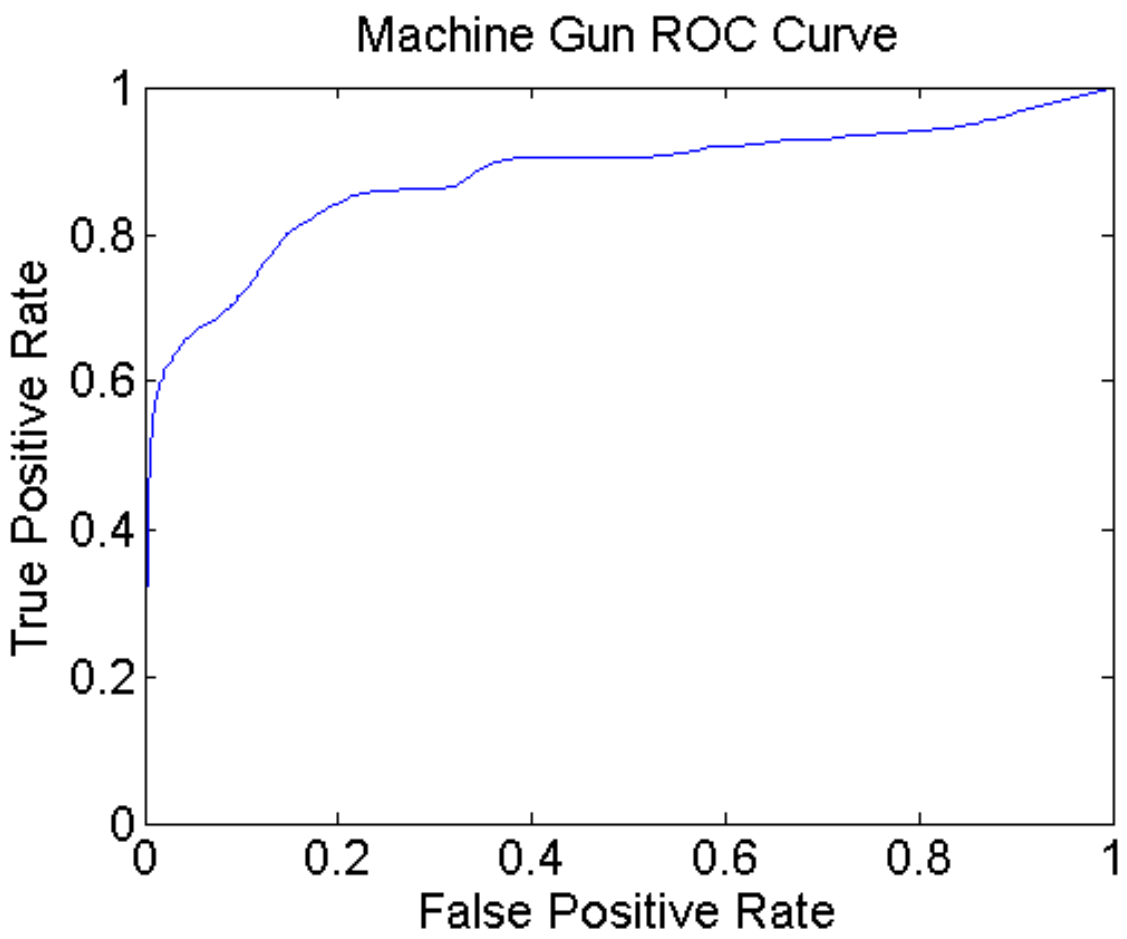

Figure 64: Machine Gun ROC Curve

Figures 65, 67, and 69 show the boxplots of ANN outputs according to vehicle, aircraft, and thunder respectively. Like machine gun, their true positive ANN outputs were less tightly clustered at the median.

Figures 66, 68, and 69 show that vehicle, aircraft, and thunder exhibit steep drop offs in true positive rates as the threshold decreases from 1; not as drastic as machine gun, but more severe than blast or wind.

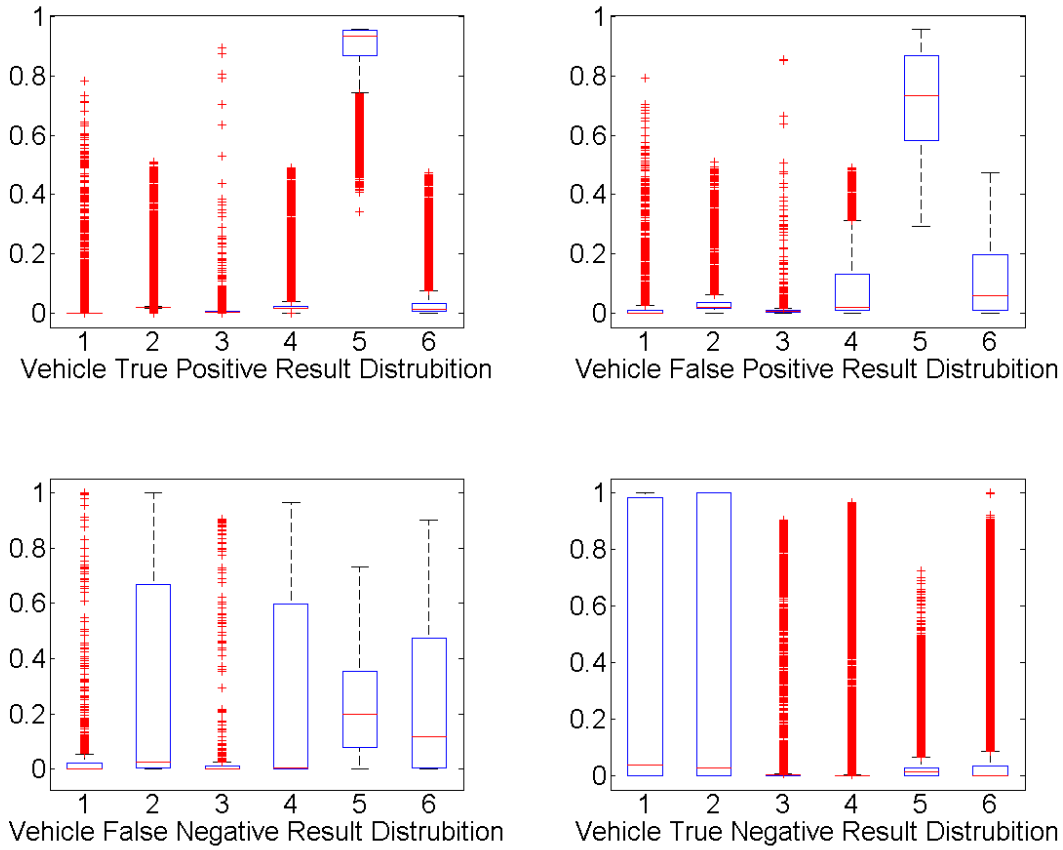

Figure 65: Vehicle Result Distribution

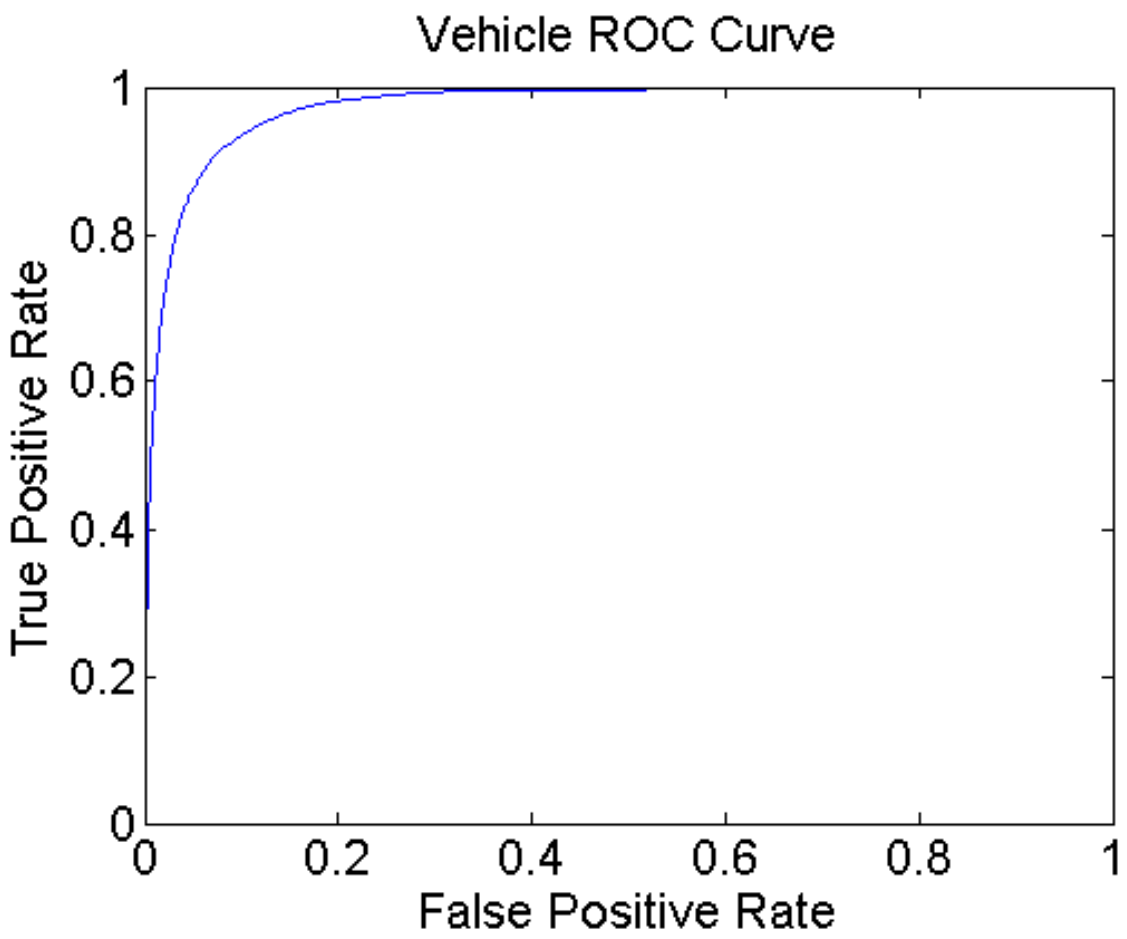

Figure 66: Vehicle ROC Curve

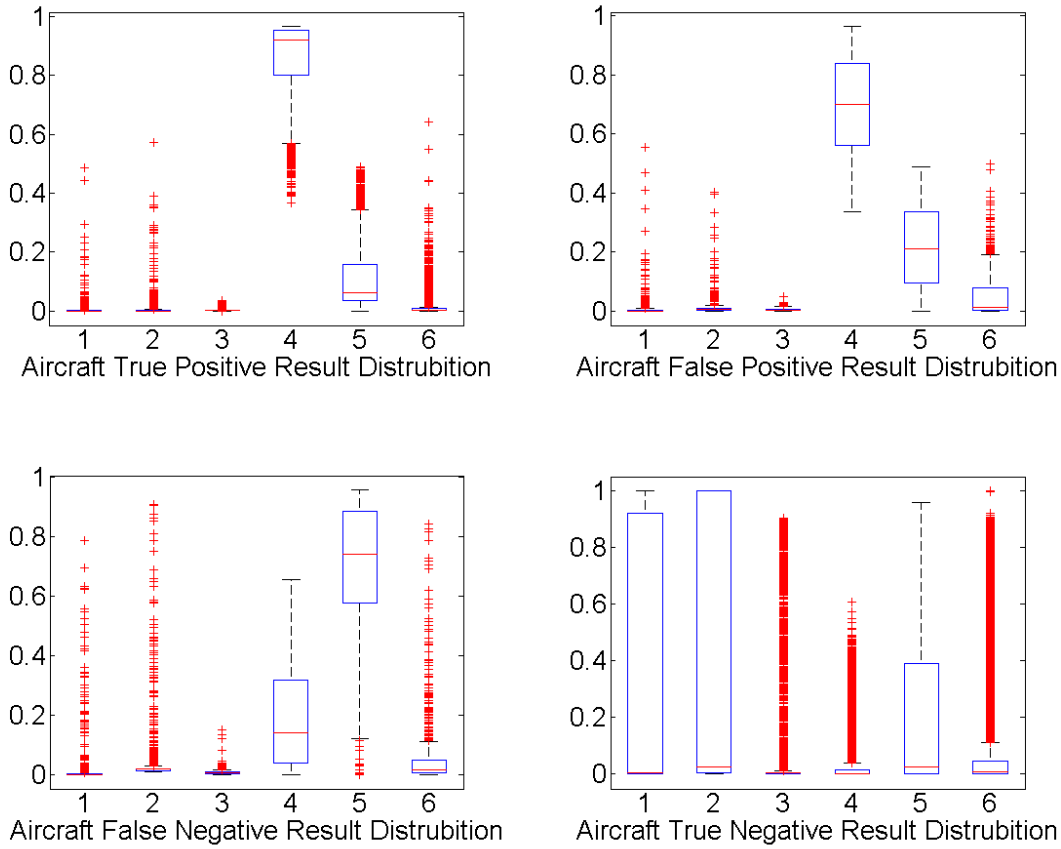

Figure 67: Aircraft Result Distribution

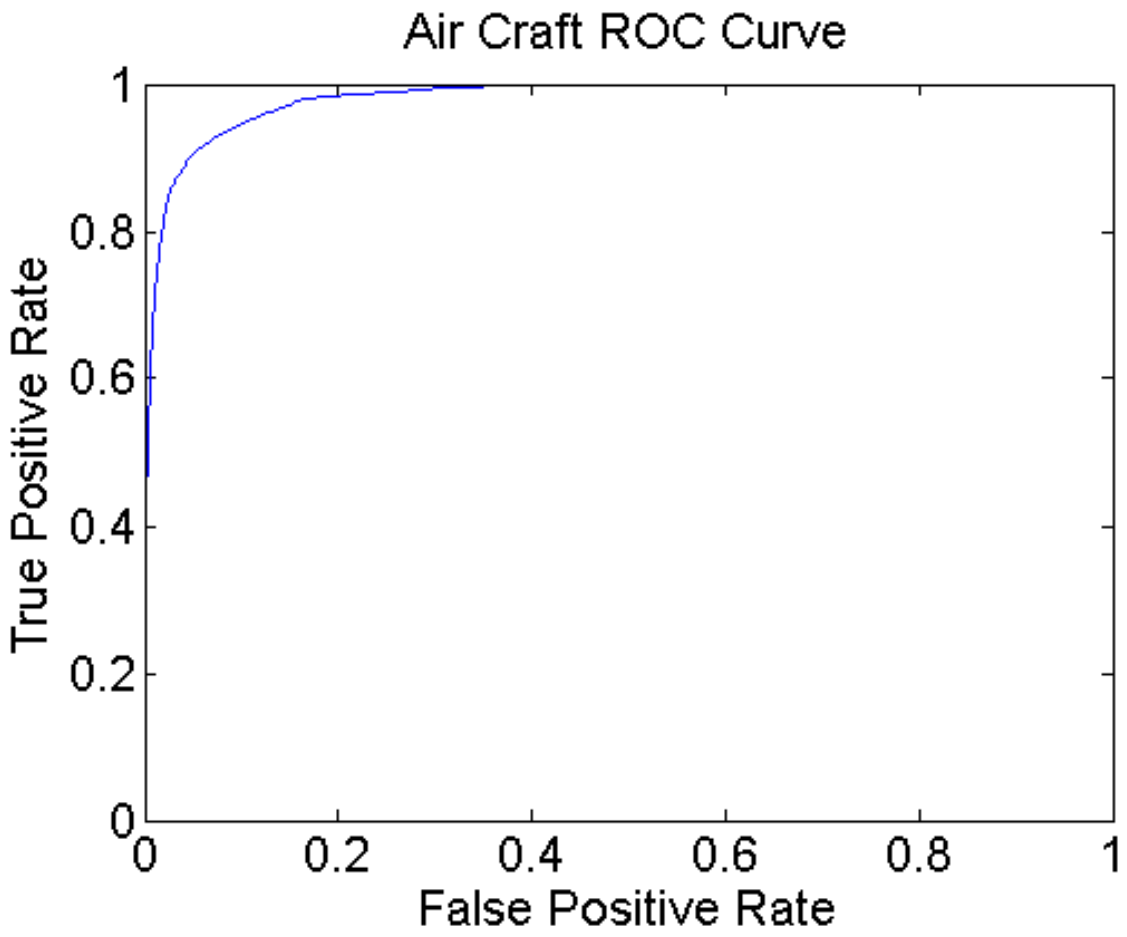

Figure 68: Aircraft ROC Curve

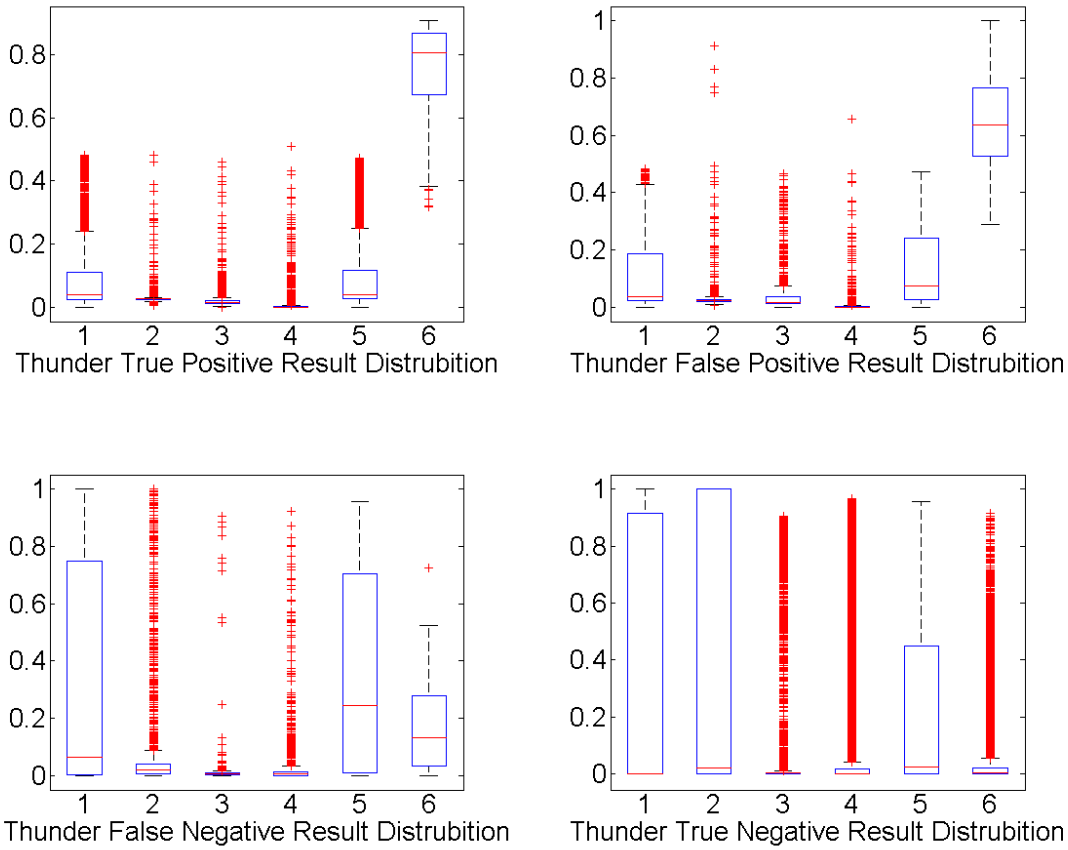

Figure 69: Thunder Result Distribution

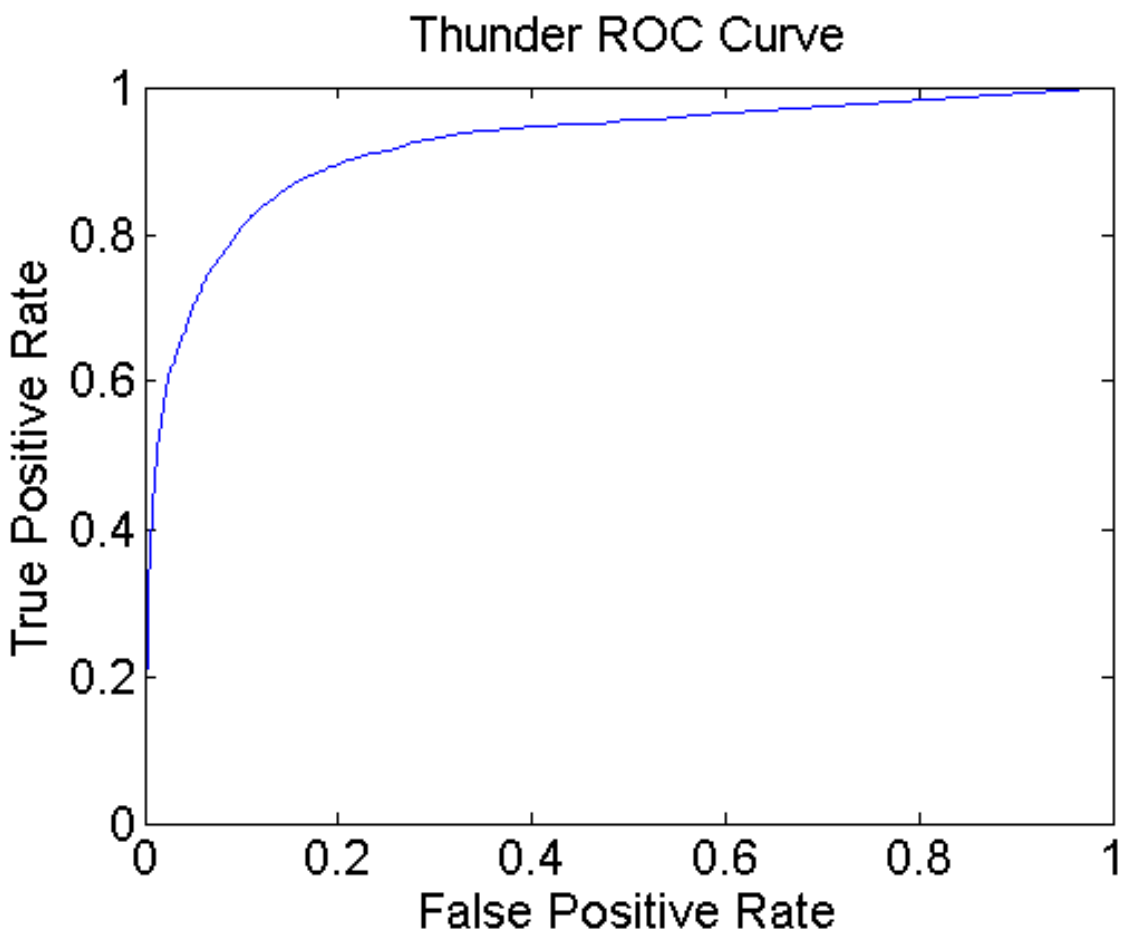

Figure 70: Thunder ROC Curve

In all cases, there is usually a large gap between the majority of true positive values in the correct class and those of the other classes, but only about 75% of all true positives fall within that gap. The high median ANN output of blast false negatives also proved that even that generalization proved inaccurate when the signal to noise ratio dipped too low. With large areas of overlap between ANN correct and incorrect output possibilities a simple adjusted threshold was determined not to be a credible method of increasing classifier accuracy.

Figures 71-76 display histograms of the second highest result for each sound class. For almost all false negatives with the exception of machine gun, the second highest result was usually correct. This begged the question as to whether or not there was a way to determine if the first or second highest ANN output was correct.

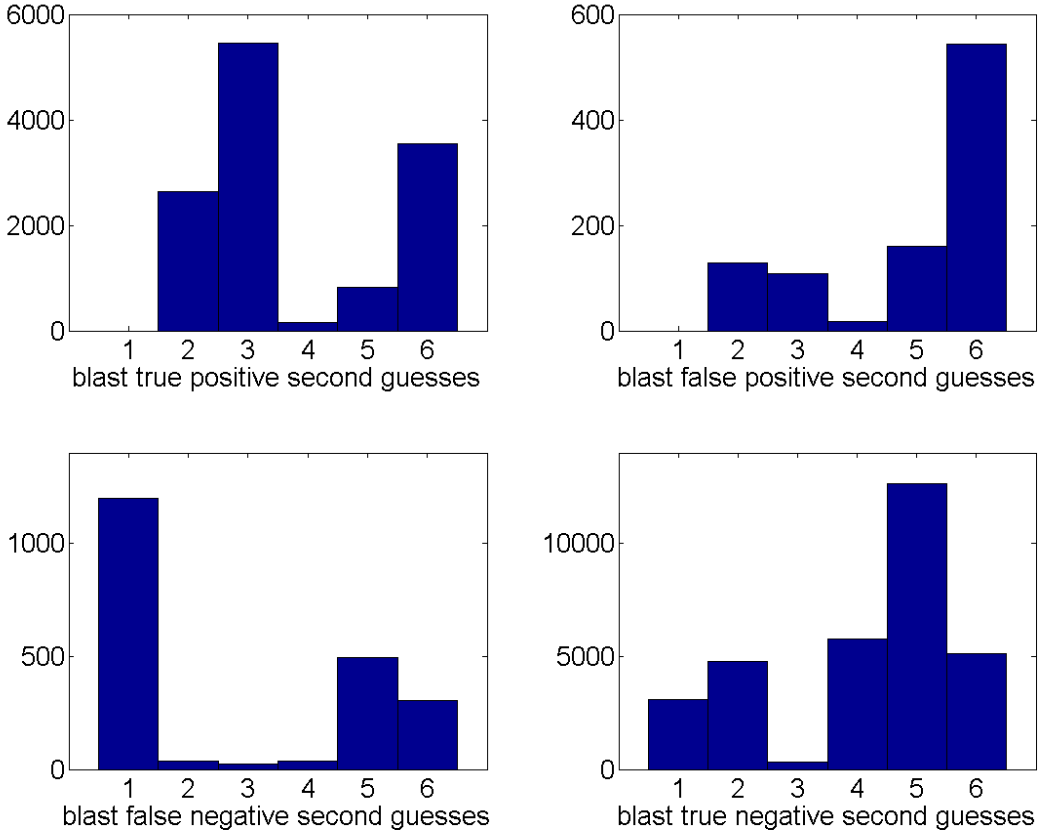

Figure 71: Blast Second Guess Histogram

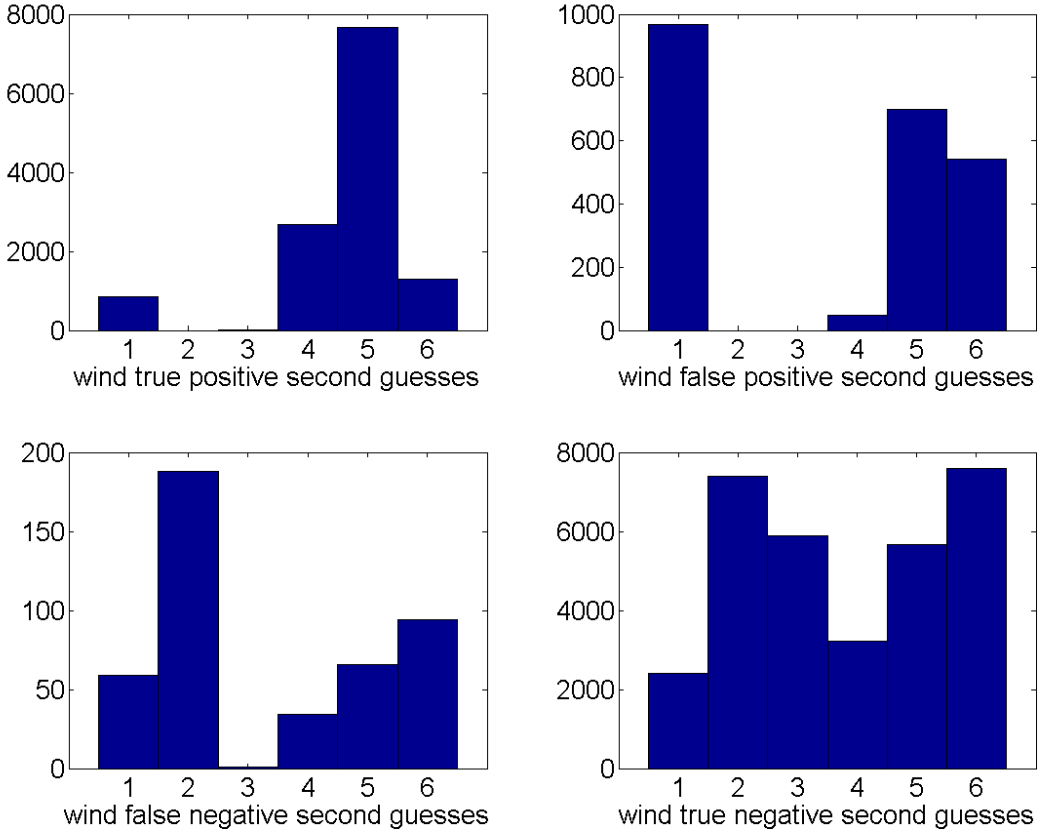

Figure 72: Wind Second Guess Histogram

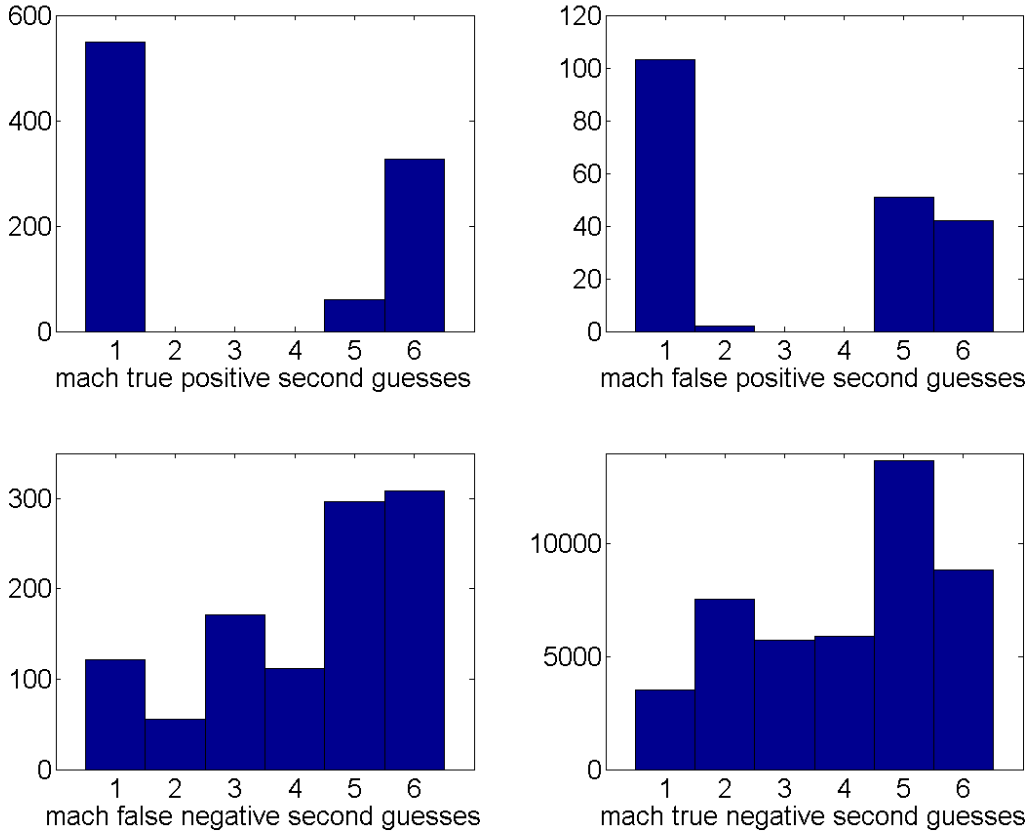

Figure 73: Machine Gun Second Guess Histogram

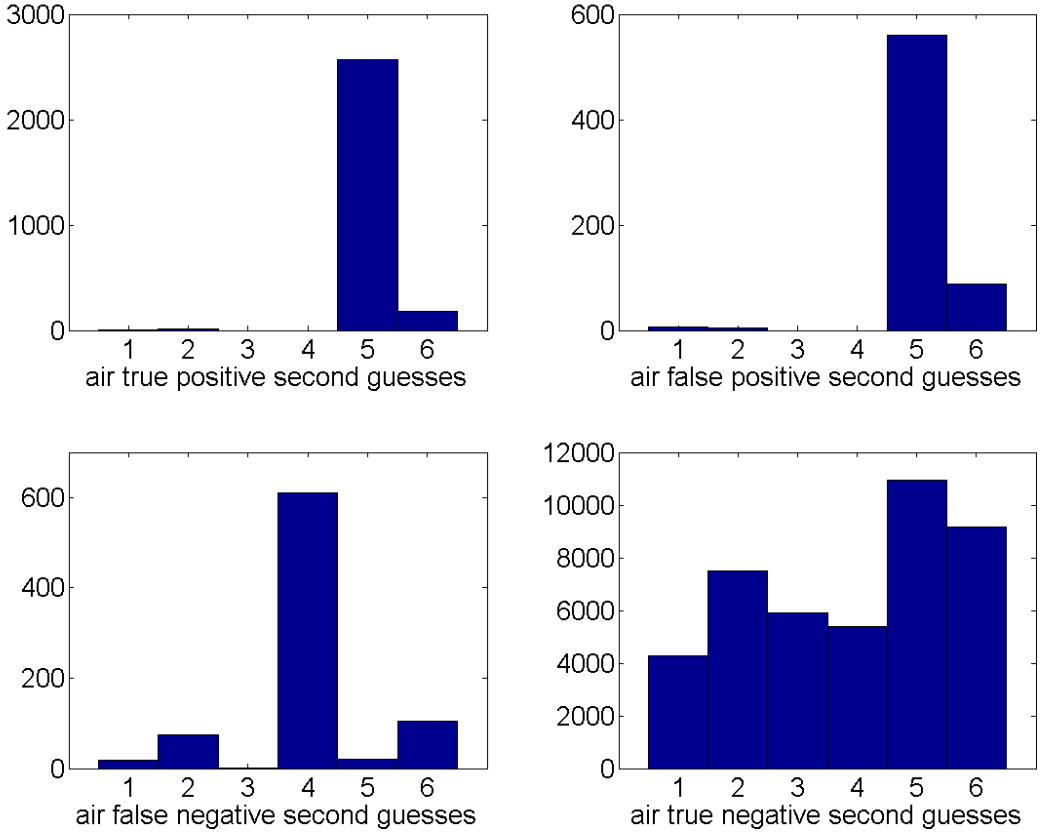

Figure 74: Aircraft Second Guess Histogram

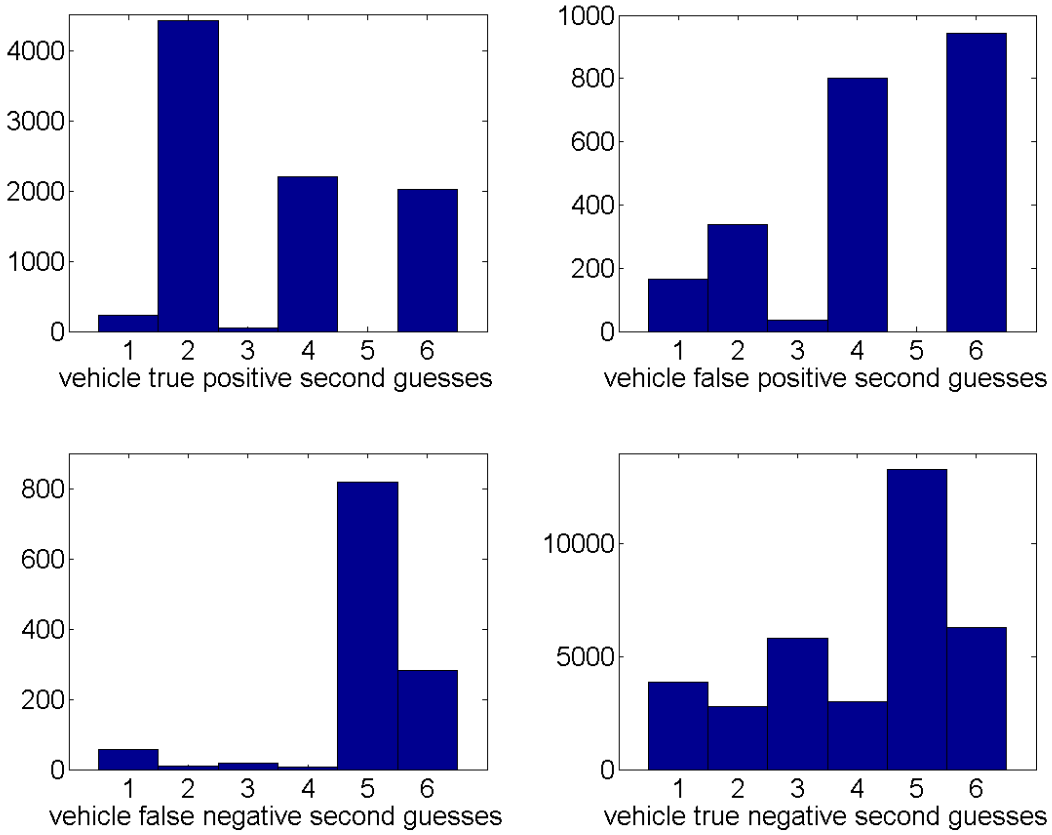

Figure 75: Vehicle Second Guess Histogram

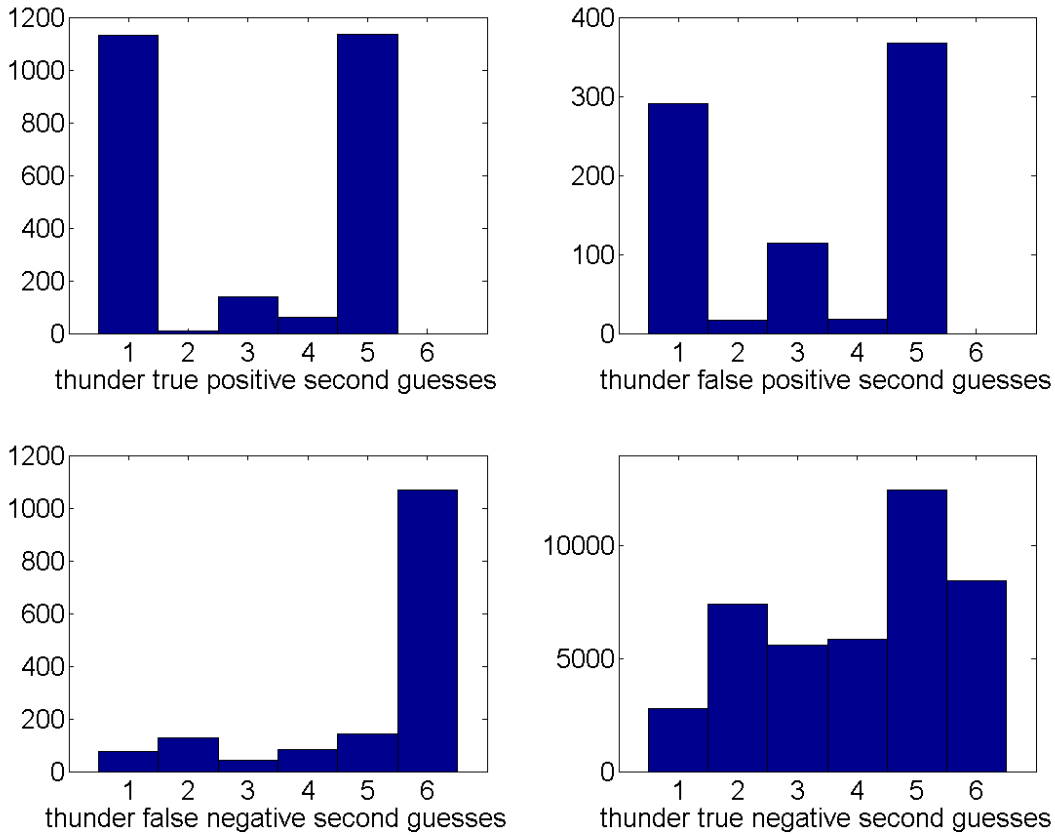

Figure 76: Thunder Second Guess Histogram

Figure 77 displays the difference between the 1st and 2nd highest ANN outputs when the highest result is correct while Figure 78 displays that same difference when the second highest ANN output is correct. Blast and wind true positives have, with a few exceptions, a very large difference between 1st and 2nd largest ANN outputs when the 1st highest is correct. Machine guns, aircraft, and vehicles follow the same trend but to a lesser extent. Thunder did not. In every case, the difference between the 1st and 2nd highest ANN outputs when the 2nd highest ANN output is correct was the same.

In summary, threshold methods of ANN evaluation are problematic for multiclass classifiers because raising any threshold increases the likelihood of null classifications and lowering thresholds increases the likelihood of dual classifications. It is possible on average to use the difference between 1st and 2nd highest ANN output when that difference is greater than 0.2 to provide confidence for blast, wind, machine gun, air craft, and vehicle true positive classifications.

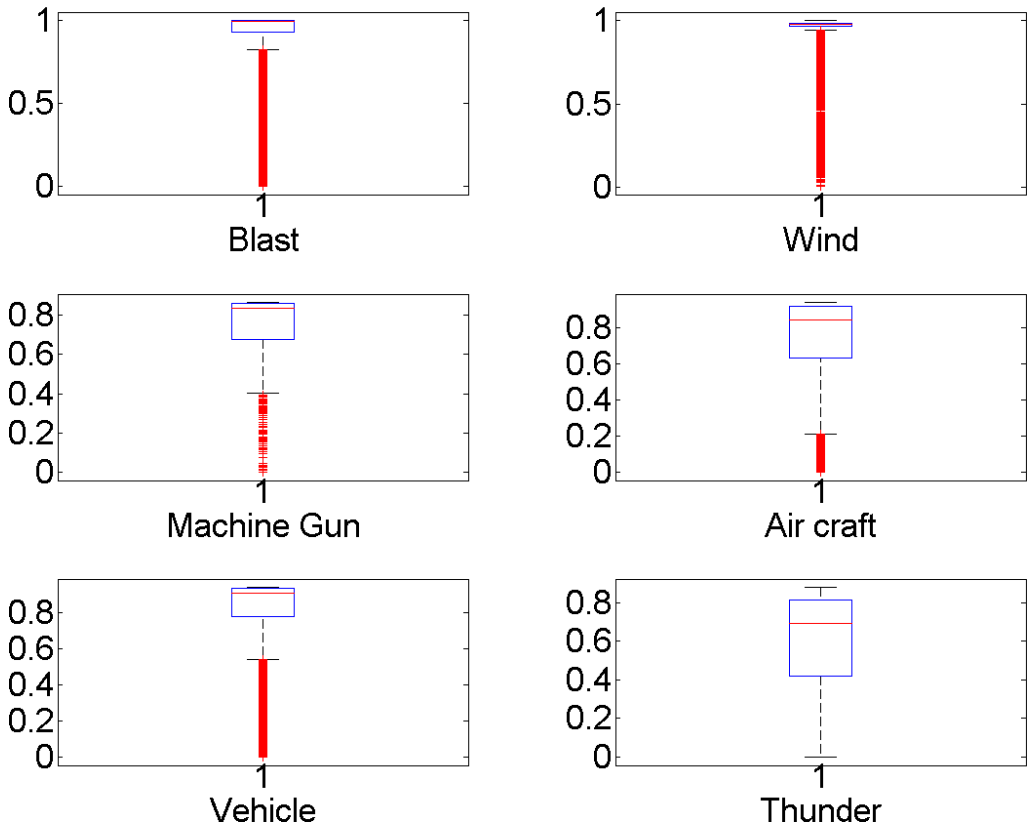

Figure 77: Difference Between the Satlin Squashed First and Second Highest ANN Output (Highest Output is Right)

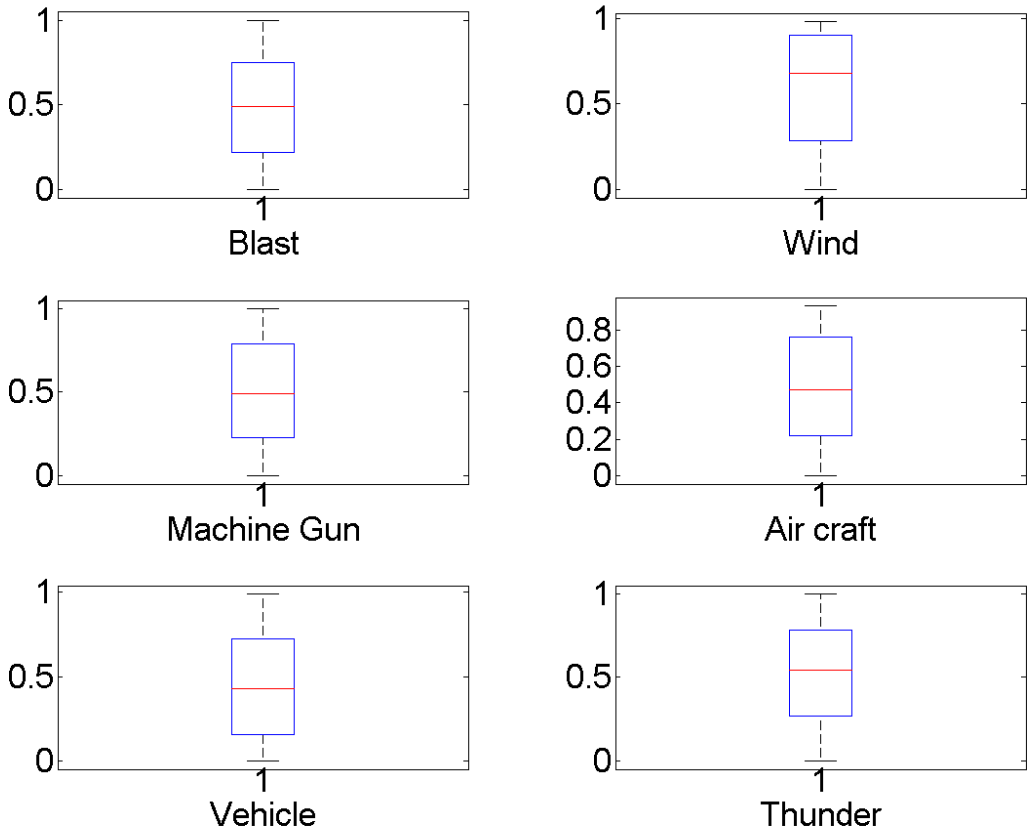

Figure 78: Difference Between the Satlin Squashed First and Second Highest ANN Output (2nd Highest Output is Right)

10.0 CONCLUSIONS

An improved military noise classifier was developed using a multilayer perceptron (MLP) artificial neural network (ANN). The previous ANN classifier could identify three (3) classes of noise: (wind, blast, and aircraft), whereas the classifier presented here can identify six (6) types of noise (wind, blast, aircraft, vehicle, electronic noise, and machine gun).

The efficacy of the original UPITT classifier was determined by a site visit where BAMAS microphone arrays were installed. There, researchers personally confirmed the source of every sound event they heard and matched them with events logged by the BAMAS system. Under windy situations when considering all files matched, the original UPITT classifier was found to have an overall accuracy of 94.9% because 94.7% of the events were accurately classified wind files. It's blast true positive rate was 2.9% but the majority (100%) were not recorded due to wind filtering. Of the files recorded, it accurately classified 90.6% of them. On a calmer day with intermittent thunderstorms, the original UPITT classifier had an overall accuracy of 75.5% but a blast true positive accuracy of 87.2%. When examining only the files which bypassed the wind filtering, it accurately classified 90.5% of them.

"Regularization" and "Early Stop" training methods were considered, but the accuracy difference between the two was minimal. The training time for the regularization method; however, was considerably longer, so all ANN's were trained using the early stop method.

Twenty one (21) new acoustic metrics were considered including linear FFT centroids, number of peaks in the time history, and a host of sound level meter metrics which boosted the overall accuracy of the classifier by 11.36%. The "round" and "max" methods of ANN evaluation were compared. In conjunction with the new metrics, the "max" method only offered a 1.83% overall accuracy increase, but resulted in no null classifications; making it the preferred method.

Due to a large amount of confusion between aircraft and vehicle sound classes, helicopters and small single engine planes were reclassified as vehicles due to their similar time histories. The result was a 3.26% increase in overall accuracy and a 64% drop in vehicle and aircraft confusion.

The inclusion of wind and mixed blasts were toggled to gauge their effect on classifier accuracy. It was found that excluding either had minimal effects on blast identification but drastically reduced the ANN's ability to properly classify wind and mixed blast files. Although the wind filtering system is supposed to prevent wind files from being recorded, should wind files be recorded they would most likely be falsely classified as blasts without proper wind training. And blast mixed with even the slightest amounts of wind would be rejected without including mixed blasts during training; resulting in an increased sensitivity to wind.

A signal to noise ratio (SNR) analysis found that the Eight FSS Metric classifier retained 97.5% accuracy at a SNR of 10 (dB) which was 23.65% higher than the Lopatka classifier and 11.00% higher than the Chan classifier. The increasing sample size (ISA) analysis showed that a database of at least 34,000 samples was needed to achieve minimum error and 47,464 samples were used during thus study's analyses.

Seven test subjects were asked to classify 1000 sound files. When the new ANN classified the same files, it scored just as accurately as the median test subject in class.

A Forward Sequential Selection revealed that only eight (8) metrics (ASEL, Kurtosis, X linear FFT centroid, ZSEL, LA max fast, A peak, spectral slope, and crest factor) were needed for the overall accuracy of the classifier to remain at 90.04%.

A wind speed analysis proved that not only was there was a correlation between wind speed and number of missed classifications, but that the new classified improved accuracy by between 4.34% and 10.36% on two separate occasions.

Network pruning showed that an ANN with three hidden layers with seven neurons each provided as much accuracy as more complex ANN configurations with an overall accuracy standard of deviation less than 1% and a mean overall accuracy of 90% guaranteeing the reliable high accuracy.

Lastly, when the output distributions of the ANN were examined it was found that there is on average a large gap between the first and second highest output values; indicating an optional metric of confidence.

The result of these training parameters and ANN structure was a new classifier using only eight (8) metrics, seven neurons for each of it's three hidden layers which proved to be more accurate than the original UPITT classifier under every circumstance. It improved blast overall accuracy by between 4% and 10.36% in windy conditions, decreased blast false positive rates by 1.66%, cut blast false negative rates by 15.9%, and increased blast true positive rate by 15.9%. The number of machine gun shots mistaken as blast decreased by 77%. Overall, it was projected to have an accuracy of 90.37%.

11.0 FUTURE WORK

The individual site ISA analyses proved that site specific training can produce high accuracy using less training data. Site specific training requires three (3) steps: installation and calibration of the equipment, sound file classification of new data from that site, and ANN updating. For most military bases, steps two and three are infeasible because trained staff would be needed to classify thousands of sound files, train a new network, and update the system.

The FSS metric classifier can classify acoustic events in real time; however, human classification is needed to verify it's classification accuracy. There were thousands of these events per month at the Base 3 installation and classifying the files alone was a daunting task for one person; however, it would have been easier for thousands of people to classify one or two files each. Google trains its map making algorithms to read addresses from blurring pictures via crowd sourcing methods. They require users to type the numbers they see into text box to confirm that they are human. A similar "human check" could be implemented where users would have to listen to a sound file and type what they believe caused the acoustic event. Because there is some variability with human classification, logic could be used to only accept a certain classification if three or five other users also selected it.

Once a crowd source method of sound classification is completed, the library of sound files for each site will slowly but surely be classified; producing a fully classified data set. Scripts can be written to train a new ANN using data gathered from that site. The new ANN's performance would be compared to the old one and if it provides an improvement this system is upgraded. This system would allow for site specific training of an ANN, but more importantly it would provide a model for future adaptive ANN's to learn the subtleties and nuances of new information from their specific location using data obtained from unconventional crowd sourcing methods.

BIBLIOGRAPHY

[1] E. Nykaza.
An investigation of community attitudes toward blast noise, general community survey, study site 1.
ERDC-CERL, TP-19-9, 2012.

[2] E. Nykaza and G. Luz.
Improved procedure for correlating blast noise events with complaint logs at u.s. army instalations.
Noise Control Engineering Journal, 56 (6), 2008.

[3] Alfonso Chacon-Rodgriguez and Pedro Julian.
Evaluation of gunshot detection algorithms.
Proceedings of the Argentine School of Micro-Nanoelectroincis, Technology and Applications, 2008.

[4] Cheung-Fat Chan and Eric W.M. Yu.
An abnormal sound detection and classification system for surveillance applications.
In *European SIgnal Processing Conference*, 2010.

[5] Kuba Lopatka.
Dangerous sound event recofnition using support vector machine classifiers.
Springer-Verlag Berlin Heidelberg, 2010.

[6] Brian Bucci.
Development of artifical neural network-based classifiers to identify military impulse noise.
Master's thesis, University of Pittsburgh, 2007.

[7] Brian A. Bucci and Jeffrey S. Vipperman.
Performance of articial neural network-based classifiers to identify military impulse noise.
Journal of the Acoustical Society of America, 122(3):1602–1610, September 2007.

[8] Robert M. Cvengros.
Blast noise classification with common sound level meter metrics.
Acoustical Society of America, page (in press), 2012.

[9] Jeffrey S. Vipperman and Brian A. Bucci.
Development and implemetation of metrics for identifying military impulse noise.
Strategic Environment Research and Development Program (SERDP), 2008.

[10] Matthew Brandon Rhudy.
Real time implementation of a military impulse noise classifier.
Master's thesis, University of Pittsburgh, 2009.

[11] Brian A. Bucci and Jeffrey S. Vipperman.
An investigation of the characteristices of a bayesian military impulse noise classifier.
ASME IMECE Conference, no. NCAD2008-73046, 2008.

[12] *Nueral Network Toolbox User Guide: Hidden Markov Models*, 1984-2010.

[13] Taylor Sauder.
The implementation of a hidden markov model in matlab for the prediction of commodity prices.
Foster College of Business Administration, Bradley University, 2011.

[14] J. S. Vipperman and B. Bucci.
Algorithm development for a real-time military noise monitor.
Final SERDP Report SI- 1436, March 24, 2006.

[15] J. S. Vipperman.
Development and implementation of metrics for identifying military impulse noise.
Final SERDP Report Draft for Project SI- 1585, June 18, 2010.

[16] B. Bucci and J. S. Vipperman.
Development of an artificial neural network-based classsifier to identify military noise.
Journal of the Acoustical Society of America, 122(3):1602–10, September 2007.

[17] J. S. Allanach.
Impulse noise bearing and amplitude measurement and analysis system.
Final SERDP Report Draft for Project SI- 1427, March 21, 2010.

[18] Applied Physical Sciences Corp.
Bamas dem/val data description.
Technical report, Applied Physical Sciences Corp., 475 Bridge Street, Suite 100, Groton, CT 06340, July 19, 2012 2013.

[19] E. Billaur.
Peak detection using MATLAB.
http://www.billauer.co.il/peakdet.html., 2011.

[20] M.P. Norton and D.G. Karczub.
Fundamentals of Noise and Vibration Analysis for Engineers.
Cambridge University Press, 2003.