Developing Predictive Athletic Performance Models for Informative Training Regimens

Jordan E. Blanchfield, Meredith T. Hargroves, Peter J. Keith, Maryanna C. Lansing, Lars Hälsing Nordin, Rachel C. Palmer, Shelby E. St. Louis, Allyson J. Will, William T. Scherer, and Nicholas J. Napoli
University of Virginia, jeb8xa, mth4aq, pjk5ry, mcl8dr, lhn2rz, rcp5vx, ses3mu, ajw7kh, wts, njn5fg@virginia.edu

Abstract - **Individualized biometric data are being incorporated into training and competitions by many coaches and trainers to provide insights into athletic performance and physical fitness of their athletes. Currently, fitness tracking software provides coaches with minimal descriptive statistics on the collected biometric data, resulting in limited actionable outcomes. The collection of biometric data provides an opportunity to understand the variables that are indicative of athletic performance, and to create predictive models to determine appropriate training and in-game strategies. In order to develop these informative decision support tools, predictive frameworks have to address the correct performance metrics, control of subject-to-subject variability, handle data limitations, and maintain model interpretability. We demonstrate that the strenuousness of training sessions leading up to a competitive match has significant impact on the outcome of the game (win or loss) in continuous-play team sports. Specifically, a high cardiovascular training load two days prior to competition was predictive of a win. Additionally, we show that statistically significant differences exist in the physiological behaviors of different player positions. Analysis of several performance metrics also demonstrates that singular metrics or combinations of simple statistics do not directly relate to the outcome of a game, particularly in low-scoring sports such as field hockey or soccer.**

Index Terms - athletic performance, biometric data, heart rate analytics, predictive modeling, sports analytics, training load

INTRODUCTION

The introduction of wearable technologies to professional and collegiate sport has potential to transform the way coaching and training staffs use data to monitor players and manage training regimens and competition strategies. These devices collect real-time biometric data, enabling deeper insights to athlete's physiological states. Teams use biometric data in many capacities. These applications include customizing training, preventing injury, and monitoring players' physiological attributes both on and off the field [1]. Statistics and predictive models can even be applied to recruiting efforts and competition strategies. However, as teams seek to understand and make use of biometric data, the need for advanced analytics and actionable modeling is rapidly increasing.

Because wearable devices are a relatively new development, coaches, trainers, and managers are unfamiliar with continuous biometric data collection, making interpretation and validation of models difficult [2]. Current analysis is mostly descriptive, and the lack of context makes it difficult to develop actionable solutions [2]. Therefore, there is demand for data infrastructures which can efficiently handle and analyze biometric player data and models that can provide actionable recommendations for both training and in-game strategy.

I. Prior Work

Monitoring fitness using biometric data is still in its infancy. Currently, the overwhelming amount of data has not been surpassed by the ability to process it in meaningful ways [3]. Some studies on important factors, including training load and various heart rate measures, have been performed successfully. Studies show a high intensity training period followed by a taper period reduces training load and allows players to fully recover for competition [4, 5, 6]. This work recognizes the complexity of team sports in terms of the influence of player positioning has on performance metrics but has failed to control for such variability [6]. Other challenges of measuring performance include the numerous factors that affect the individual and team performance including tactical, physiological, social, and external components [7, 8]. It is clear univariate modeling is unlikely to predict changes in a multifaceted outcome accurately, but the number of attempts using advanced multivariate approaches is extremely limited [7]. This raises three main questions: (1) How can we control for the variability across players and positions when modeling and predicting performance? (2) How can we account for the multifaceted nature of performance? (3) What are strong predictive features in terms of individual player and team performance?

II. Challenges

Team sport performance is a complex and multidimensional construct, and different positions on the team require various skills and demand different amounts and types of physical output during games [7]. Thus, developing a model around such high variability would greatly influence the model

978-1-7281-0998-5/19/$31.00©2019 IEEE

accuracy and interpretability. Team sport play can generally be categorized as either discrete or continuous. In continuous-play sports, such as field hockey, soccer, and many others, the lack of discrete events creates difficulty in modeling because there are very few single, comprehensive metrics of individual performance [9]. Logical individual performance metrics include statistics such as goals, assists, and other sport specific metrics, such as circle penetrations or ball wins in field hockey. However, game outcomes vary significantly when it comes to these performance metrics. There is no obvious relationship between a single performance metric and a team win or loss.

Because wearable technology is relatively new and constantly evolving, limited prior research is available. Many algorithms used by wearable companies are proprietary, and there is very little exchange and communication among analysts [10]. Thus, modeling player performance is dependent on the generation of novel biometric features that explain the physiological state of a player.

III. Insights

Player positions are so inherently different, thus it is not reasonable to assess them together in a single model. Variability across players can be controlled for by stratifying the population by player position to gain a better understanding of how players operate under different demands. This is a crucial aspect of the modeling process that prior work has failed to recognize and enables more training samples than individual models. To account for multifaceted dimensions of performance, multiple game statistics such as the number of circle penetrations and the number of shots taken on the goal can be grouped into a single performance metric using unsupervised learning techniques. It is believed that unsupervised clustering can efficiently and accurately handle all the multifaceted performance metrics in the data to better model human performance. Modeling these responses by training on engineered biometric features specific to player position could potentially provide accurate informative predictive models to coaching and training staff.

IV. Contributions

This analysis of biometric data in continuous-play sport leads to three key contributions:

1. Significant statistical differences exist between physiological metrics for different player positions.

2. The outcome of a game can derive from several different combinations of performance statistics, and single metrics do not directly result in a win or loss.

3. The combination of Training Loads one (TL 1), two (TL 2), three (TL 3), and four (TL 4) days before a game affects both individual and team performance.

Methods

I. Data Collection

The University of Virginia's (UVA) varsity field hockey team collected data to address these research questions for examining player differences, performance metrics and predictive modeling. The Polar monitors track biometric data during practices and games. Additional data sources provided information such as game statistics (e.g., score, circle penetrations), hydration, and weather conditions. A proprietary, longitudinal data frame combines these features, in addition to biometric features engineered from the 10Hz Polar data streams (heart rate, distance, velocity, acceleration, cadence, GPS, R-R intervals) for each player.

II. Data Infrastructure

Each Polar data file synchronizes with the start and end time of each half of each game using velocity and GPS data for starting players. Patterns in the velocity and GPS location data indicate the start and end times since players participate in a pre-game meeting and other predictable routines.

Matlab 2018B code imported the external data relating to individual players or specific games. This external data includes the number of minutes a player played in the game, the low and high temperature for the day, the rating percentage index (RPI) of the opponent, the distance the team traveled, and many other stats provided by the coaching staff or found online.

The imported 10 Hz biometric Polar data files use timestamps from zero rather than time of day. Synchronizing the data with the start and end times of the game halves requires parsing the filename to obtain the session start, and adding it to the data timestamps. Referencing the new timestamps against the start and end time of each game ensures players wore the monitor for the entire duration of the game.

Complex features are engineered as proxies for athlete preparedness and fitness and incorporated into the data frame that combines sport-specific statistics (i.e. goals scored, circle penetrations), external statistics (i.e. outside temperature, opponent rating percentage index), and player biometric data (i.e. heart rate variability, training load). A function creates three rows, which correspond to the first half of the game, the second half, and the game as a whole, containing player metrics and engineered features. This is repeated for every player and game, resulting in a condensed dataset used for analysis.

III. Modeling

Predictive models are trained on the longitudinal data frame that contain player-level metrics for each half of the game. Multifaceted response variables quantify performance, which cannot be captured in a single game statistic. Candidate metrics include ball wins and losses, circle penetrations, shots on goal and goals. However, each player position has unique game demands that lead to differing performance metrics; for this reason, unique response variables quantify performance for each position.

Predictive features from the longitudinal data set were selected in combinations and trained to predict performance using supervised learning techniques including trees. The models used both raw features and clusters of features. The nodes of the most accurate models determine which features and interactions of features contribute to performance. Additional models including support vector machine (SVM) classification used these features as predictors to improve accuracy.

Data Analysis and Results

I. (RQ 1) What statistical differences are there across position?

Visual inspection of boxplots for several physical measurements confirm hypotheses that there are differences in the biometric demands of offensive, midfield, and defensive players. Kruskal-Wallis tests verify the statistical significance of these differences for several features such as Average Heart Rate, Heart Rate Range, Standard Deviation of RR intervals, and others. Kruskal-Wallis tests allow different groups to be compared in non-parametric distributions with the null hypothesis that the respective means of each group is equal. Table 1 illustrates these differences for Mean values by position and Kruskal-Wallis results.

TABLE I
MEAN VALUES BY POSITION AND KRUSKAL-WALLIS RESULTS

Features	Offense		Midfield		Defense		p-value
	μ	SE	μ	SE	μ	SE	
Average Heart Rate	156.0	1.7	162.1	2.0	164.1	2.0	0.0159
Heart Rate Range	115.6	4.7	143.9	6.3	111.4	4.9	0.0029
Standard Deviation of Heart Rate	24.1	0.5	26.4	1.0	20.5	1.0	< 0.001
Standard Deviation of RR Intervals	0.1	0.0	0.2	0.0	0.1	0.0	0.0212
Permutation Entropy of RR Intervals	1.3	0.0	1.2	0.1	1.3	0.1	< 0.001

Due to these observed statistical differences between positions, further analysis was conducted on datasets stratified by position.

II. (RQ 2) How do we characterize multifaceted performance metrics?

Existing knowledge and collaboration with UVA field hockey faculty confirmed that field hockey is a relatively low-scoring game with very few discrete events that serve as obvious performance metrics. Due to the continuous nature of the game, the final game outcome is not always indicative of each player's contribution. As seen in Figure I, the result of a game (represented in either orange or blue) can derive from several different combinations of secondary performance statistics.

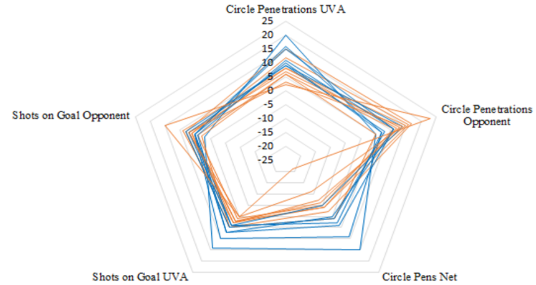

FIGURE I
PERFORMANCE METRICS RADAR CHART

For example, in one game last season, UVA won by seven goals and the opposing team only had five circle penetrations (an indicator of strong defensive performance by UVA). However, in another game, the opposing team won by two goals, also with five circle penetrations. The variability in secondary performance metrics and limited insight available from the final outcome demonstrate the need for innovative response variables that capture the multimodal nature of the game.

In order to capture position-specific demands, the stratified datasets were augmented with logical, multifaceted response variables. For example, appropriate metrics to evaluate the offensive players are the number of circle penetrations and the number of shots taken on the opponent's goal. Defensive players, however, cannot be assessed with the same metrics since the responsibilities of the position are different. Given the inability to quantify optimal human performance with a single number, an alternative way to characterize performance is through unsupervised k-means clustering, which produces discretized states. Through an iterative process manipulating k values and observing related centroids, response clusters were selected based on two criteria: 1) distinguishable clusters with a clear optimal state and 2) relative class balance. Table 2 shows offensive response states. Cluster 2 is the strongest state with a clear set of desirable centroids, corresponding to optimal secondary performance metrics.

TABLE II
OFFENSIVE PERFORMANCE CLUSTER

	Circle Pen's (UVA)	Shots on Goal (UVA)	Goal Diff	Aggregate Ball Net	N
Cluster 1	5	1	-2.8	-7.4	14
Cluster 2	20	12.6	4.7	-1	11
Cluster 3	10.8	4.4	0.2	-1.3	35
Cluster 4	10.6	5.4	-0.7	-10.5	19

These performance clusters were used as response variables in predictive modeling.

III. (RQ 3) How does predictive modeling of performance inform training decisions?

Since response variables were defined as discrete states, classification trees were a logical modeling technique. Various combinations of physiological features were trained on fine, coarse, and medium tree models using a 5-fold cross validation. Cross validation is a valid method for nonparametric model selection and is useful when datasets are limited. Additionally, it provides a way to test how well the model will generalize to new data [11].

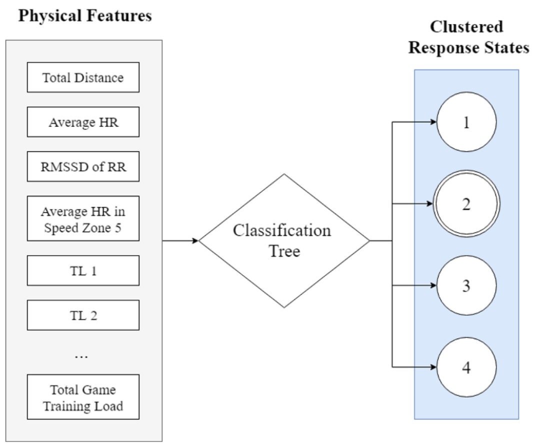

FIGURE II
VISUAL REPRESENTATION OF PREDICTIVE MODELING

The model predicting offensive response clusters yielded a classification tree comprising of five features: TL 1, TL 2, Proportion in Speed Zone 5, Average Heart Rate in Speed Zone 5, and Standard Deviation of Heart Rate. Figure III shows the confusion matrix of the true positive rates for the classification states. State 2 had a true positive rate of 73%, however there were only 11 subjects in that class, so the projected confidence interval for accuracy ranges from 64% to 82%.

FIGURE III
CONFUSION MATRIX OF TRUE POSITIVE RATES

This model provides valuable insight through the interpretation of the resulting classification tree. Figure IV displays part of the tree, with the black dot signifying the optimal offensive performance state.

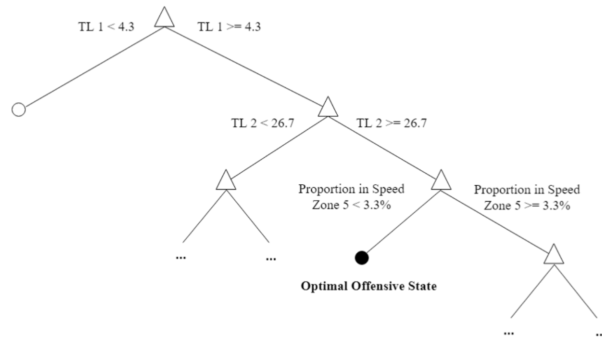

FIGURE IV
OFFENSIVE CLASSIFICATION TREE

From this tree, it is evident that TL 1and TL 2 Days Prior are the biggest drivers of state 2. The tree suggests that TL 2 should be above 26.7, which is above the 25th quantile but below the median Training Load for all practices two days before competition. This means that the offense performs better when TL 2 is at a relatively medium or high level. Although the response variable in this model was a stratified state designed for offensive players, we hypothesized that strong offensive performance would correspond to overall game outcome. Graphical inspection of TL 1 versus TL 2 for all player positions confirmed the hypothesis that the thresholds determined in our classification tree can also model game outcome, as depicted in Figure VI.

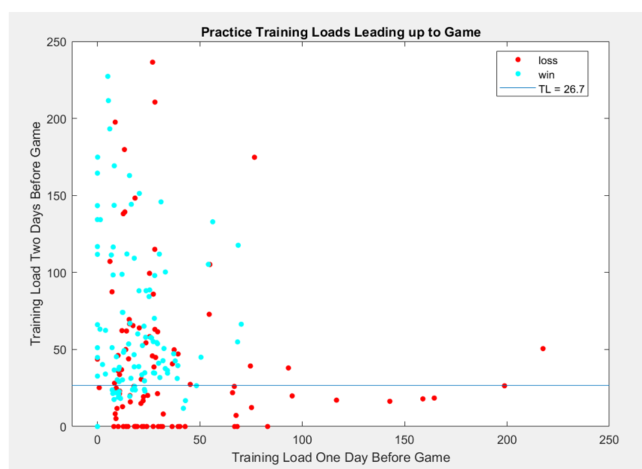

FIGURE V
TL1 VS. TL2 BY GAME OUTCOME

As shown in the horizontal line in Figure V, this decision boundary creates a model that is able to classify game outcome with 72.8% accuracy. The strong results from the use of a single threshold suggested that the use of a SVM, with more complex decision boundaries, could improve the overall accuracy of our model. Although initial investigation only included TL 1 and TL 2, an SVM is able to handle multi-dimensional predictor variables and thus, all four days of Training Load data were included in the new model. Figure VI provides a visual explanation of the process. Unlike previous models classifying unique states of performance, the response variable in the SVM model is a binary loss (0) or win (1).

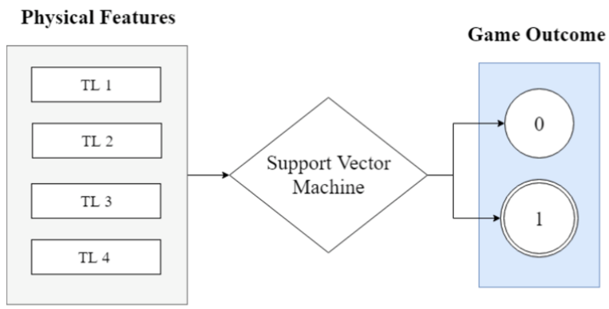

FIGURE VI
VISUALIZATION OF PREDICTIVE MODELING

The results of our trained SVM yield 79.8% accuracy in classifying game outcome from Training Load from all four days prior to competition. The strong performance of our classification model using only Training Load data suggests that the strenuousness of practices leading up to a competition can be highly influential in determining the outcome of a game. The Training Load data that was used to train this model was continuous data; however, the unique numerical value of each player's Training Load is less significant than the relative magnitude of each value compared day-to-day.

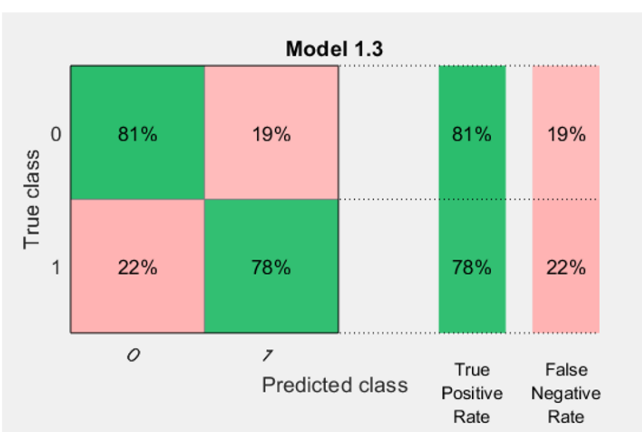

FIGURE VII
CONFUSION MATRIX FOR SVM MODEL ACCURACY OF THE MODEL: 79.8%.

For this reason, the Training Load that results from a certain practice can be considered as one of three levels: high, medium, or low. With this generalization, the set of all possible combinations from four days of practice prior to a game results in 81 unique patterns (i.e. HMML, HLMH, etc.). Given the accuracy of our SVM in classifying game outcome, this trained model can be applied to the 81 unique combinations of Training Load data and predict the associated game outcome. Although these results cannot be validated, the predictions for each unique pattern serve as a method of validation for recommendations from earlier stages of exploration and modeling.

CONCLUSION & FUTURE WORK

The results of our research derive from several investigations and validation methods. Using a fairly limited sample of offensive field hockey players, Training Load Two Days Prior to a game emerged as a highly discriminatory feature in predicting whether or not a player would reach the optimal performance state. Specifically, conducting a medium or high-intensity practice two days before a competition frequently led to strong offensive performance. The hypothesis that this threshold is significant for all players was confirmed by extrapolating this rule as a new model for predicting overall game outcome, which was ultimately successful with 72.8% accuracy for every player, regardless of position. Using 228 observations of players across all positions, a SVM classified game outcome with 79.8% accuracy. SVM classification increased the complexity of the decision boundary between predicting a game win or loss and allowed for a new test set of hypothetical Training Load data to be classified with the trained model. Finally, the output of the model predictions can be combined with early insights on Training Load (derived from classification trees with position-specific responses) to validate our hypothesis about conducting medium or high-intensity practices two days before a game. Predictions from our trained SVM classify 100% of patterns with "low" TL 2 as a loss. Therefore, the research team recommends that practice plans two days prior to competition are designed to reach medium or high-intensity in order to avoid low Training Load values.

The accuracy and complexity of the models involved in this research vary due to the different classification problems that each model is tasked with. Although the best-performing model does not incorporate multifaceted physical data, restricting the response variable to a binary assignment significantly reduces the complexity of the model. Although the accuracy of the four-state classification is lower than the binary-classification, the four-state classification attempting to predict unique performance states is significantly more complex. Therefore, the tree classification models used with multifaceted features still provide valuable insights moving forward in analysis.

These preliminary results provide insights into the optimal difficulty of field hockey practices in the days leading up to a game, but include several restrictions. Since data from only one season was used in the predictive models, the results are limited, but provide a foundation for future work and analysis of continuous sports with multifaceted indicators of player performance. The models were trained on a maximum of a couple hundred rows of data, creating concern over the confidence interval on the accuracy. The success of our models with limited data is encouraging and suggest that this methodology can lead to actionable insights as more data is collected in the future.

The lack of response metrics recorded by halves of games limited potential insights into in-game substitution schedules and strategies. In order to aid subjective in-game coaching decisions, coaches should record metrics such as ball wins and losses, shots on goals, and goals by each player at the half level. These response metrics would make analysis possible at a more granular level than the entire game, allowing insights into the players' state changes from one half to the next and providing indicators of fatigue.

Future development will include improvements to the predictive models for defensive and midfield players in order to understand the differences in features that influence performance for players that have different in-game demands. Additional analysis may focus on modeling specific events within games such as ball losses or circle penetrations to understand the physiological state that occurs in a player immediately before a successful play or unsuccessful action. These models would enable real-time analysis with the potential to provide in-game insights to the coaches.

Acknowledgements

We would like to thank the UVA Field Hockey Team and coaching staff for allowing us the opportunity to work with them during the 2018-2019 school year. A special thank you goes to Rachel Dawson for helping make this project happen and being the capstone team's main point of contact on the athletics side.

References

[1] K. Karkazis, and J. Fishman, "Tracking U.S. professional athletes: The ethics of biometric technologies." The American Journal of Bioethics, vol. 17, no. 1, 2017.

[2] L. Saxon, "Athletic performance monitoring, pseudo science and metaphysics meet ethics," The American Journal of Bioethics, vol. 17, no. 1, pp. 61–62, 2017.

[3] Emerging airXv, "Big data analysis is changing the nature of sports science," MIT Technology Review, March 2016 https://www.technologyreview.com/s/600957/big-data-analysis-is-changing-the-nature-of-sports-science/. Accessed: March 2019.

[4] N.J. Napoli, M.W. Demas, S. Mendu, C.L. Stephens, K.D. Kennedy, A.R. Harrivel, R.E. Bailey, L.E. Barnes, "Uncertainty in heart rate complexity metrics caused by R-peak perturbations," Computers in Biology and Medicine, vol. 103, pp. 198-207, 2018.

[5] V. Manzi, S. D'Ottavio, F. Impellizzeri, A. Chaouachi, K. Chamari, and C. Castagna, "Profile of weekly training load in elite male professional basketball players," The Journal of Strength & Conditioning Research, vol. 24, no. 5, 2010.

[6] I. Mujika, "Intense training: the key to optimal performance before and during the taper," Scandinavian Journal of Medicine & Science in Sports, vol. 20, September 2010.

[7] C. Schneider, F. Hanakam, T. Wiewelhove, A. Döweling, M. Kellmann, T. Meyer, M. Pfeiffer, A. Ferrauti, "Heart rate monitoring in team sports—A conceptual framework for contextualizing heart rate measures for training and recovery prescription," Frontiers in Physiology, vol. 9, May 2018.

[8] N.J. Napoli, L.E. Barnes, ``A Dempster-Shafer Approach for Corrupted Electrocardiograms Signals''. AAAI Proc. of the 29th Int'l Florida Artificial Intelligence Research Society (FLAIRS) Conference. (2016), Key Largo, FL.

[9] W. A. Sands, A. A. Kavanaugh, S. R. Murray, J. R. Mcneal, and M. Jemni, "Modern techniques and technologies applied to training and performance monitoring," International Journal of Sports Physiology and Performance, vol. 12, no. 2, 2017.

[10] J. Schrack, V. Zipunnikov, and C. Crainiceanu, "Electronic devices and applications to track physical activity," JAMA, vol. 313, no. 20, pp.2079–2080, May 2015.

[11] P. Zhang, "Model selection via multifold cross validation," The Annals of Statistics, vol. 21, no. 1, March 1993.

Author Information

Jordan E. Blanchfield, Meredith T. Hargroves, Peter J. Keith, Maryanna C. Lansing, Lars Hälsing Nordin, Rachel C. Palmer, Shelby E. St. Louis, Allyson J. Will, William T. Scherer: Department of Engineering Systems and Environment, University of Virginia

Nicholas J. Napoli: Department of Electrical and Computer Engineering, University of Virginia