An Ensemble Machine Learning Model for the Prediction of Danger Zones: Towards a Global Counter-Terrorism

**Abstract:** Terrorism is a global menace which has stayed with humanity since the ancient times. It is a global concern – a feeling that hypothetically cause people to make more conservative and less risky decisions often as a way of compensating for the feelings of insecurity caused by the disaster. The inclinations and impacts of terrorist activities are often quantified and assessed by the number of incidents and casualties. Terrorism is counter-terrorism; hence its negative impacts on global economy. There is therefore need to apply sophisticated computational models to combat global terrorism. Much research efforts have been dissipated in the area of counter-terrorism but fewer efforts are placed on terrorism prediction. The aim of this work is to develop a stack ensemble machine learning model which combines Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) for the prediction of continents in which a terrorism type can occur. The danger of terrorism to lives and properties in a global sense and the need to cub it is a good justification for this work. Data was obtained from an online repository known as the Global Terrorism Database (GTD). Data preprocessing included data cleaning and dimensionality reduction. Two feature selection techniques, Chi-squared (CS) and Information Gain (IG) including a hybrid of both (HB) were applied to the dataset on which a stack ensemble model was later applied. CS-reduced, IG-reduced and HB-reduced datasets produced an accuracy of 94.17%, 97.34% and 97.81% respectively at predicting danger zones with respective sensitivity scores of 82.3%, 88.7% and 92.2% and specificity scores of 98%, 90.5% and 99.67% respectively. These imply that the HB technique produced the best accuracy in least time in predicting the location of a possible occurrence of terrorism incidents. This work has established major elements of a terrorism incident that are peculiar to each of the six continents of the world. This adds to what was known about the Global Terrorism Database. This will help in making informed decision on tourism and global counterterrorism.

**Keywords:** terrorism, danger zones, tourism, ensemble machine learning, support vector machine.

**INTRODUCTION**

Terrorism is a global menace which has stayed with humanity since the ancient times. It is a global concern – a feeling that hypothetically cause people to make more conservative and less risky decisions often as a way of compensating for the feelings of insecurity caused by the disaster (Sacco, Galletto, & Blanzieri, 2003). One of the most popular events in time is the 9/11 which has not only been recorded as one of the deadliest single attack in history but also has attracted world’s attention to the dire need of investigating, predicting and curbing its far-reaching global effects. Although terrorism has often been defined in a way that is specific to the subject-matter of the particular convention, it is described by Title 22 of the U. S. code as politically motivated violence perpetrated in a clandestine manner against non-combatants (Roser, Nagdy, & Ritchie, 2019). It is often committed so as to create a fearful state of mind in an audience very often different from the victims. Although some have argued that terrorism has some positive implications, it cannot be legitimate even if it was later discovered to be benign (Sorel, 2003). This was premised on the fact that any action exhibited against a representative public order is an illicit act and can be characterized as oppressive and illegitimate. This is true because terrorism itself is associated with violence, extremism, intimidation and acts against public and social order.

The danger of terrorism to lives and properties in a global sense and the need to cub it is a good justification for this work. The inclinations and impacts of terrorist activities are often quantified and assessed by the number of incidents and casualties (Frey, Luechinger, & Stutzer, 2007). Frey et al. noted that terrorists can be empirically proven to often succeed in inflicting massive civilian casualties and colossal damage. Example is the case of attacks on the World Trade Towers in the 9/11 and the bombings of three railway stations in Madrid in March 11, 2004. It was further ascertained (Gunaratna, 2008) that terrorist attacks vis-à-vis anti-colonial movements, ethnic separatist groups amongst others all try to harm their opponents in order to force them to yield to their demands or, more, to overwhelm them.

Martha Crenshaw argued that terrorism is a logical choice which happens when the power ratio of government to aggrieved is high (Crenshaw, 1981). Crenshaw categorized the causes of terrorism into three layers: the situational factor, the strategic factors and the individual factors. The situational factors include conditions that allow the possibility of radicalization and motivate feelings against the enemy as well as specific triggers for actions.
The strategic factors may, in the long run, mean political change, nationalists, and revolution or separatist movements. It may mean, in the short run, an act to advertise a cause. It may also seek to disrupt and discredit the process of government, influence public attitude and prevent good governance, instill fear and sympathy as well as provoke a counter-reaction to legitimize their grievances. The individual factors deal with the worldview, psychology and character traits of terrorists. It assumes that terrorist personality or predisposition exists in humans.

Several research works have been done in the application of statistical, machine learning and deep learning techniques to the Global Terrorism Database (GTD) towards countering global terrorisms. Some authors have applied machine learning techniques such as Naïve Bayes (NB), K-Nearest Neighbors (KNN), Decision Trees and its variants and Support Vector Machine (SVM) to predict the terrorist groups responsible for a given incident (M. Khorshid, T. Abou-El-Enien, & G. Soliman, 2015; Tolan & Soliman, 2015). Some authors have also developed a recommender system using deep learning to predict the rate at which terrorists spread online propaganda, although the performance of this model was not evaluated (Alhamdani, Abdullah, & Sattar, 2018). A hazard grading model has also been developed for the quantification of terrorist attacks using K-Means clustering (Jun, Tong, Zhiyi, & Yutong, 2019). The success of terrorist activities has also been predicted using Decision Tree Algorithms (Jani, 2019). Some researchers also predicted if a particular terrorist attack is targeted at a government official, civilians, military, business or others(Salem & Naouali, 2016). Similar work was also done in (M. M. Khorshid, T. H. Abou-El-Enien, & G. M. Soliman, 2015), (Nieves & Cruz, 2011) and (Adnan & Rafi, 2015). Whether a given terrorist attack will be claimed by a known group or not has also been established by recent researchers using different machine learning techniques (Kumar, Mazzara, Messina, & Lee, 2018). Unfortunately, none of these works has predicted location, that is, continents where terrorisms can occur. Such information can aid the War on Terrorism; improve security consciousness and serves as good advice for tourists.

The aim of this work is to develop a computational model for the prediction of danger zones as it relates to terrorism.

Materials

In this study, we obtained our dataset from the University of Maryland’s online repository known as the Global Terrorism Database (GTD). The data is maintained by the National Consortium for the Study of Terrorism and Responses to Terrorism (START). The data description shows that an incident is categorized and recorded as terrorism if it meets any two of these three criteria: (1) it is intentional, (2) it entails some level of violence or immediate threat of violence and (3) the perpetrators of the violence must be sub-national actors. This dataset contains incidents of terrorism and attacks that were collected from news sources all over the world (LaFree & Dugan, 2007). The GTD contains 181,691 instances of recorded incidents of terrorism attacks recorded from July 1970 to December 2017. The dataset has 139 attributes, many of which are sparse due to missing data. Detailed description of this dataset is available elsewhere (Dugan, 2011; LaFree, 2010; LaFree & Dugan, 2007). In order to give our model a statistical power and to reduce fitting error, we removed all the attacks that do not have complete information required for the modeling.
METHODS

We developed a workflow for the study to guide the project execution. This workflow includes preprocessing, feature selection, training, testing and prediction (Figure 1).

Figure 1: Workflow of Proposed Model

**Data Preprocessing**

This is a fundamental step in the Knowledge Discovery (KD) process. Every raw dataset collected from real-world scenarios has a high chance of ambiguity, impreciseness and inconsistency. This is often referred to as “noise” in the dataset. In this study, the preprocessing done includes data cleaning, data discretization and data normalization.

Data cleaning involves filling in missing field, identifying outliers and handling of inconsistencies in the dataset. Data discretization is an important data preprocessing task. Data discretization was done by converting the nominal fields in the GTD dataset into corresponding numerical values.

Data normalization, also known as data transformation, is vital for modeling purposes. The values in the GTD dataset are adjusted to a range. In this work, the Min-Max normalization technique was used. Min-Max is a normalization strategy which linearly transforms \( x \) to a new value \( z \) as given in Equation (1).

\[
z = \frac{x - \text{min}(x)}{\text{max}(x) - \text{min}(x)}
\]

Equation (1)

where \( \text{min} \) and \( \text{max} \) are the respective minimum and maximum values in variable (feature) \( x \) given its range. The values of a numeric range of dataset are reduced to a scale between 0 and 1.

The original GTD dataset contains 181,691 instances of terrorist attacks with 139 features of raw data, some of which contains sparsely distributed fields. Columns with 20% or less of missing data were removed. This reduced the dimension to 181691 x 46. Columns which describe similar features were removed and only one of such was retained and some extracted features replaced a feature subset. For instances, we removed the column for event ID and retained day, month and year of occurrence. Also, continent code was generated using longitude and latitude.

**Feature Selection Process**

This is the process of identifying the most relevant predictive attributes in a dataset. This will help in the identification of the feature set that are deemed very relevant for predicting continents of a particular terrorism incident. This will contribute positively to the performance and reduce the complexity of our predictive models as will be verified in the next chapter of this work.

Two filter-based feature selection methods (Chi-Square and Mutual Information Gain) were used in this study to determine the “best fit” features in the cleaned GTD datasets. This will:
i. define relevance by identifying the attributes that are more correlated to the target class (i.e. continent column)

ii. Reduce computation time because less features are usually relatively computationally less expensive.

Chi Squared and Mutual Information Gain helped in the individual ranking of the GTD features from which we selected features with higher scores. Hybrid selection was done by selecting the intersection of both selection techniques.

Ensemble Models Development for the GTD Dataset

Following data preprocessing and feature selection, we proceeded to the development of a predictive model using the qualified variables. Our prediction model is a mapping function which consists of the training dataset S of the original features. If n datasets are selected randomly for training the models using the known relevant attributes, the mapping ξ: X_j,k → Y_k defined as ξ(X_j,k) = Y_k ∀ terrorism incidents, k; where X_j,k are the set of attributes, j is the incident counter Y is the output - predicted class. Ensemble models are preferred to single models in order to reduce bias and increase the predictive power of the developed system. The motivation for using ensemble models is to reduce the generalization error of the prediction (Diop, Emad, Winter, & Hilia, 2019). As long as the base models (SVM and KNN) are diverse and independent, the prediction error decreases when the ensemble approach is used. This approach models the wisdom of crowds in making a prediction. Even though the ensemble model has multiple base models within the model, it acts and performs as a single model. KNN and SVM were chosen for this stack ensembling because they are fast and performs relatively less complex computations in predicting outcomes in a large dataset (Demidova, Klyueva, & Pylkin, 2019; Weston et al., 2001).

K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a supervised machine learning algorithm whose major idea is to learn the outcomes of the training set and predict new instances on the basis of the labels of its closest neighbor in the training set. Regardless of the largeness of the training set in the domain, new input sets can be predicted faster than other predictive models. When given an unknown tuple without its associated output class, KNN classifier searches the pattern space for the K training tuples that are closest to the unknown tuple. K in KNN is the number of the closest neighbors in the Euclidean distance. The Euclidean distance between two tuples X_1 and X_2 is given in Equation (2).

\[
\text{dist}(X_1, X_2) = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2}
\]

In the GTD dataset used in this study, the training sets are mapped to the value of the testing test and the K closest neighbors to the new outcome will vote for the predicted value and the highest vote wins.

Support Vector Machine

For our GTD dataset, given the training dataset \( \theta \) containing feature vectors \( x_i \) and their respective class labels, \( y_i \), in the form of \( \theta = \{(x_1, x_2, \ldots, x_n)\} | x_i \in \mathbb{R}^n \) where \( m \) is the dimension of the feature vector and \( n \) is the number of instances in the dataset. In the SVM technique, optimal margin classification is achieved by finding a hyperplane in \( m \)-dimensional space. The linear classifier is based on the Equation (3).

\[
f(x) = \sum_{i=1}^{n}(w_i * x_i + b) \]

where the vector, \( w_i \), is the weight vector and \( b \) is the hyperplane bias and if \( f(x) = 0 \) is called the hyperplane.

The choice of the value of \( K \) was based on independent experiments of the odd values in the interval \([3, 9]\) and optimum outcome was achieved when \( K \) was set to 3. In the SVM model, radial basis function (RBF) was used because it has localized and finite response along the entire x-axis and performed better than the sigmoid function in this experiment.

Performance Metrics

The entire study can broadly be divided into the preprocessing and prediction phases. However, we must evaluate each of these phases to ascertain our confidence in the predictive models we are developing. We used cross-validation to measure our preprocessing stage and confusion matrix as parameters for our prediction phase.
RESULTS

The data preprocessing reduced the number of features in the dataset to 21 as presented in Table.

| SN | Attribute                      | Description                                                                 |
|----|--------------------------------|-----------------------------------------------------------------------------|
| 1. | Month of occurrence            | This field contains the number of the month in which the incident occurred. In the case of incident(s) occurring over an extended period, the field will record the month when the incident was initiated. |
| 2. | Day of occurrence              | This field contains the numeric day of the month on which the incident occurred. In the case of incident(s) occurring over an extended period, the field will record the day when the incident was initiated. |
| 3. | Region of occurrence           | This field identifies the region in which the incident occurred. The regions are divided into 12 categories. |
| 4. | Specificity of Location        | This field identifies the geospatial resolution of the latitude and longitude fields. The most specific resolution uniformly available throughout the dataset is the center of the city, village, or town in which the attack occurred. |
| 5. | Is_Criteria1_Met               | These variables record which of the inclusion criteria (in addition to the necessary criteria) are met. |
| 6. | Is_Criteria2_Met               |                                                                              |
| 7. | Is_Criteria3_Met               |                                                                              |
| 8. | Is_event clearly a terrorism   | It is Boolean showing if terrorism criteria is well satisfied or satisfied. In those cases where several attacks are connected, but where the various actions do not constitute a single incident (either the time of occurrence of incidents or their locations are discontinuous, then “Yes” is selected to denote that the particular attack was part of a “multiple” incident. |
| 9. | connected_to other Attacks     |                                                                              |
| 10.| If Suicide                     | This variable is coded “Yes” in those cases where there is evidence that the perpetrator did not intend to escape from the attack alive. |
| 11.| Type of Attack                 | This field captures the general method of attack and often reflects the broad class of tactics used. It consists of nine categories, which are well defined. |
| 12.| Target Type                    | If the attacked person or group is connected to a business firm, government establishments, police, military, airports, etc. |
| 13.| Nationality of Target          | This is the nationality of the target that was attacked, and is not necessarily the same as the country in which the incident occurred, although in most cases it is. For hijacking incidents, the nationality of the plane is recorded and not that of the passengers. |
| 14.| Group Name Known               | This records if we are certain or not about the perpetrators’ group names. |
| 15.| Not Affiliated to Group        | This variable indicates whether or not the attack was carried out by an individual or several individuals not known to be affiliated with a group or organization. T |
| 16.| Type of Weapon                 | Up to four weapon types are recorded for each incident. |
| 17.| Number of Fatalities           | This field stores the number of total confirmed fatalities for the incident. The number includes all victims and attackers who died as a direct result of the incident. “Yes” appears if there is evidence of property damage from the incident. |
| 18.| Property Damaged               |                                                                              |
| 19.| If Victims in Hostage          | This field records whether or not the victims were taken hostage (i.e. held against their will) or kidnapped (i.e. held against their will and taken to another location) during an incident. |
| 20.| Ransom Given                   | If ransom was gotten for kidnapping cases |
| 21.| If Successful                  | It the attack was successful |
The longitude and latitude of each instance was also converted into corresponding continents. The resultant six (6) continents were encoded as in Table 2.

Table 2: Distribution of GTD Target Class

| S/N | Continent Code | Alpha-Encoding | Class Encoding | Number of Occurrence |
|-----|----------------|----------------|----------------|----------------------|
| 1.  | North America  | NA             | 0              | 13800                |
| 2.  | Asia           | AS             | 1              | 101788               |
| 3.  | Europe         | EU             | 2              | 21662                |
| 4.  | South America  | SA             | 3              | 18978                |
| 5.  | Africa         | AF             | 4              | 25181                |
| 6.  | Oceania        | OC             | 5              | 282                  |

The seventh continent, Antarctica, was not listed because no instance of it exists in the GTD.

The features selected from the hybridized feature selection process were collated and passed to the implemented ensemble classifier. The result of the prediction is presented in the confusion matrix in Table 3.

Table 3: Confusion Matrix of Prediction on Hybrid Feature Selected Dataset

| CLASS LABELS | Predicted | Actual |
|--------------|-----------|--------|
|              | NA        | AS     | EU     | SA     | AF     | OC     |
| N            | 440       | 29     | 61     | 1      | 0      |
| A            | 3         |        |        |        |        |
| AS           | 14        | 3340   | 139    | 78     | 159    | 2      |
| E            | 1         | 179    | 674    | 0      | 73     | 0      |
| U            |           | 7      |        |        |        |
| SA           | 61        | 48     | 3      | 610    | 0      | 0      |
| A            | 5         | 354    | 67     | 0      | 7920   | 3      |
| F            |           |        |        |        |        |
| O            | 0         | 9      | 0      | 0      | 25     | 68     |

Note: NA → North America, AS → Asia, EU → Europe, SA → South America, AF → Africa, OC → Oceania

Table 4 shows the performance of the model for the individual continent class that was predicted and the Receiver’s Operating Characteristics / Area under Curve (ROC/AUC) for the hybridized approach. This shows that the ROC for the classes ranged from 0.83 to 0.99 with class 0 (North America) having the highest area and class 5 (Oceania) having the lowest.

Table 4: Performance result for each continent

| Class | Sensitivity | Specificity | AUC  |
|-------|-------------|-------------|------|
| 0     | 0.98        | 0.98        | 0.99 |
| 1     | 0.98        | 0.96        | 0.98 |
| 2     | 0.96        | 0.98        | 0.98 |
| 3     | 0.98        | 0.92        | 0.99 |
| 4     | 0.95        | 0.97        | 0.97 |
| 5     | 0.67        | 0.93        | 0.83 |
Table 5 shows the summary of our model’s experimentation. Our model was best in terms of sensitivity, specificity and execution time when the hybrid approach was implemented.

| FEATURES            | METRICS | SCORE |
|---------------------|---------|-------|
| Chi-Square-Selection| Accuracy| 94.17 |
|                     | Precision| 93.17 |
|                     | F1-Score  | 85.33 |
|                     | Sensitivity | 82.33 |
|                     | Specificity | 98.33 |
|                     | Execution time | 83.59 |
| Info-Gain-Selection | Accuracy | 97.34 |
|                     | Precision | 96.59 |
|                     | F1-Score   | 92.83 |
|                     | Sensitivity | 88.67 |
|                     | Specificity | 90.50 |
|                     | Execution time | 84.47 |
| Combined Selection  | Accuracy | 97.81 |
|                     | Precision | 96.83 |
|                     | F1-Score   | 94.33 |
|                     | Sensitivity | 92.17 |
|                     | Specificity | 99.63 |
|                     | Execution time | 60.13 |

**Discussion**

The focus of this work was to predict terrorism as peculiar to a given continent using a stack ensemble model using a hybrid feature selection approach. The hybrid feature selection, which is the selection from both Chi-Square and Mutual Information Gain approach, has proven to be better both in the accuracies and execution time. The stack ensemble model was used for predicting continents where a particular incident is likely to occur and our feature selection techniques have yielded a high accuracy of 97.8%. Results presented in Table 5 show that the hybridized approach performs best which suggests that the proposed model is very promising at identifying locations where a given kind of terrorism incidents can occur. The ROC/AUC table shows the performance of the developed models in the light of their respective sensitivity and specificity. It is also noticed that the accuracy of class 5 (Oceania continent) has low prediction because of the support value – the total number of instances available for training and testing for the given class. This has caused a sharp deflection in the ROC for class 5.

The main strength of this work is the use of machine learning in the prediction of location in which a given terrorism incident can occur. Previous works on the GTD have investigated features such as the extent of damage of attacks, likelihood of success of a terrorism attempt, likely targets of a terrorism incident and groups likely to perpetrate an attack. To the best of our knowledge, none of the literature cited has worked on predicting the location of a terrorist attack. One weakness identified in this study is that continent was predicted even though we believed that predicting countries will be better. However, this was avoided because this will amount to too many categories and machine learning models will fail in this regard.

**Conclusion and Future Work**

In this study, we have established that the havocs wreaked by terrorist groups globally have called for a serious academic and non-academic concern. We have also achieved a tremendous result in the accuracies of the developed classifiers. The hybrid selection technique has helped in identifying seven features unique to a given continent, namely, region of occurrence, nationality of target groups, target type, property vandalism, number of casualties and type of weapon. Findings in this study imply that the HB technique produced the best accuracy in least time in predicting the location of a possible occurrence of terrorism incidents. This work has established major elements of a terrorism incident that are peculiar to each of the six continents of the world. This adds to what was known about the Global Terrorism Database and will help in making informed decision on tourism and global counterterrorism.

**References**

1. Adnan, M., & Rafi, M. (2015). Extracting patterns from Global Terrorist Dataset (GTD) Using Co-Clustering approach. *Journal of Independent Studies and Research, 13*(1), 7.
2. Alhamdani, R. S., Abdullah, M. N., & Sattar, I. A. (2018). Recommender System for Global Terrorist Database Based on Deep Learning. *International Journal of Machine Learning and Computing, 8*(6).
3. Crenshaw, M. (1981). The causes of terrorism. *Comparative politics, 13*(4), 379-399.
4. Demidova, L., Klyueva, I., & Pylikin, A. (2019). Hybrid Approach to Improving the Results of the SVM Classification Using the Random Forest Algorithm. *Procedia Computer Science, 150*, 455-461.
5. Diop, A., Emad, N., Winter, T., & Hilia, M. (2019). Design of an Ensemble Learning Behavior Anomaly Detection Framework. *International Journal of Computer and Information Engineering, 13*(10), 551-559.

6. Dugan, L. (2011). The making of the Global Terrorism Database and its applicability to studying the life cycles of terrorist. *The SAGE handbook of criminological research methods, 175.*

7. Frey, B. S., Luechinger, S., & Stutzer, A. (2007). Calculating tragedy: Assessing the costs of terrorism. *Journal of Economic Surveys, 21*(1), 1-24.

8. Gunaratna, R. (2008). Bruce Hoffman: Inside Terrorism: (New York, Columbia University Press, 2006): Taylor & Francis.

9. Jani, T. (2019). Predicting success of global terrorist.

10. Jun, Y., Tong, X., Zhiyi, H., & Yutong, L. (2019). Hazard Grading Model of Terrorist Attack Based on Machine Learning.

11. Khorshid, M., Abou-El-Enien, T., & Soliman, G. (2015). A comparison among support vector machine and other machine learning classification algorithms. *IPASJ International Journal of Computer Science, 3*(5), 26-35.

12. Khorshid, M. M., Abou-El-Enien, T. H., & Soliman, G. M. (2015). Hybrid Classification Algorithms For Terrorism Prediction in Middle East and North Africa. *International Journal of Emerging Trends & Technology in Computer Science, 4*(3), 23-29.

13. Kumar, V., Mazzara, M., Messina, A., & Lee, J. (2018). A Conjoint Application of Data Mining Techniques for Analysis of Global Terrorist Attacks. Paper presented at the International Conference in Software Engineering for Defence Applications.

14. LaFree, G. (2010). The global terrorism database (GTD) accomplishments and challenges. *Perspectives on Terrorism, 4*(1), 24-46.

15. LaFree, G., & Dugan, L. (2007). Introducing the global terrorism database. *Terrorism and Political Violence, 19*(2), 181-204.

16. Nieves, S., & Cruz, A. (2011). *Finding Patterns of Terrorist Groups in Iraq: A Knowledge Discovery Analysis.* Paper presented at the Ninth LACCEI Latin American and Caribbean Conference (LACCEI’2011), Engineering for a Smart Planet, Innovation, Information Technology and Computational Tools for Sustainable Development., Medellin.

17. Roser, M., Nagdy, M., & Ritchie, H. (2019). Terrorism. OurWorldInData https://ourworldindata.org/terrorism

18. Sacco, K., Galletto, V., & Blanzieri, E. (2003). How has the 9/11 terrorist attack influenced decision making? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 17*(9), 1113-1127.

19. Salem, S. B., & Naouali, S. (2016). Pattern recognition approach in multidimensional databases: application to the global terrorism database. *International Journal of Advanced Computer Science and Applications (IJACSA), 7*(8).

20. Sorel, J. M. (2003). Some questions about the definition of terrorism and the fight against its financing. *European Journal of International Law, 14*(2), 365-378.

21. Tolan, G. M., & Soliman, O. S. (2015). An experimental study of classification algorithms for terrorism prediction. *International Journal of Knowledge Engineering-IACSIT, 1*(2), 107-112.

22. Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., & Vapnik, V. (2001). *Feature selection for SVMs.* Paper presented at the Advances in neural information processing systems.