GiCoMAF: An Artificial Intelligence Algorithm to Utilize Maps for Operators of Unmanned Aerial Vehicles

Yuval Zak, Yisrael Parmet and Tal Oron-Gilad
Ben-Gurion University of the Negev
Beersheva, Israel
{zaky@post.bgu.ac.il, iparmet@bgu.ac.il, orontal@bgu.ac.il}

Abstract

Unmanned Aerial Vehicles (UAV) operators must maintain high levels of situation awareness on their area of operation. To achieve this, they use the Command and control (C2) map, which are shared among forces, and is regularly overloaded with data that is irrelevant to their mission. UAV operators' missions require distilled information at the right timing. Yet, the existing filtering mechanisms of C2 maps are layer-based and insufficient. We propose a new approach to automatically and dynamically filter information items on the map based on environmental and mission context. To achieve this, we introduce a three-tier artificial intelligence (AI)-based algorithm (GiCoMAF), where we delineate the use of machine learning (ML) models to support UAV missions. For the GiCoMAF development, tagged data was collected in simulated experimental runs with professional UAS operators. Different types of ML models were evaluated and fitted into the algorithm. The models achieved a relatively high accuracy at modeling human preference and area of interest. The approach presented in this study can be further implemented to support other operators in time-critical spatial-temporal problems.

1 Introduction

February 2010, Afghanistan, an American helicopter fired on three suspected trucks, killing 23 innocent civilians, and wounding 12. The attack was approved based on information provided by operators of an unmanned aerial vehicle (UAV) who did not report the presence of civilians in the trucks (Filkins 2010). In a later news report, Shanker and Richtel (2011) cite Army officials claiming that the leading cause for the tragic incident was information overload that lead to poor Situation Awareness (SA). SA can be defined as one's perception of the environment around him or her at any given point in time (Endsley 1988). Thus, although there were evidences for children in the trucks, the UAV operators "did not adequately focus on them amid the swirl of data".

The use of UAVs in the military domain is increasing, due to their ability to perform missions without risking human operators (Izzetoglu et al. 2015). UAV operators monitor the payload, often a camera sending video feed, in various missions (e.g., reconnaissance, guidance of forces; Marusich et al. 2016), in addition to multiple tasks (e.g., navigation and orientation, flying the vehicle, radio communication; Everaerts 2008). A command and control (C2) map, often in a different display, is used for orienting and making sense of the payload's outputs. The C2 map shows mission critical information and intelligence-related information items such as markings of potential targets, location of allied forces and so forth. A cognitive work analysis of UAV operators workflows, emphasizes frequent and continuous use of the C2 map during missions (Back et al. 2019). The C2 map, however, being shared among military elements, is showcasing information that is irrelevant to the UAV operators. It has been indicated (Endsley 2000; Sandom 2000) that information overload is a contributor to poor Situation Awareness high workload and low overall performance. According to Back et al. (2019), the information clutter in C2 maps may often lead operators to neglect the map, and rely solely on the payload's feed, and may also lead to fatal results as the tragic incident of February 2010. Answering Adams' (2015) call for incorporating human factors limitations in the design of UAVs, it serves as an incentive to address the information overload problem. Some advanced solutions for improving UAV operators' SA had been suggested, e.g. using synthetic vision (Calhoun et al. 2005). Such solutions, however, may require costly adjustments of the vehicle's payload.

For decluttering, most C2 maps are based on information layers. Layers can manually or automatically (via a set of rules) be hidden, shown or dimmed. The layer mechanism reduces the information overload problem by hiding or dimming layers, yet, at the same time, it may cause information deprivation due to an inherent tradeoff; any action performed on a layer affects it entirely. Thus, it is impossible to hide irrelevant information items in a layer, while showing relevant ones of the same layer. Therefore, aspiring to solve the C2 map information overload tradeoff, Zak, Oron-Gilad, and Parmet (2018) called for 'breaking' the layer mechanism and addressing the information at the information items' level. Thus, instead of filtering layers of information items, the filter will handle each information item individually (as

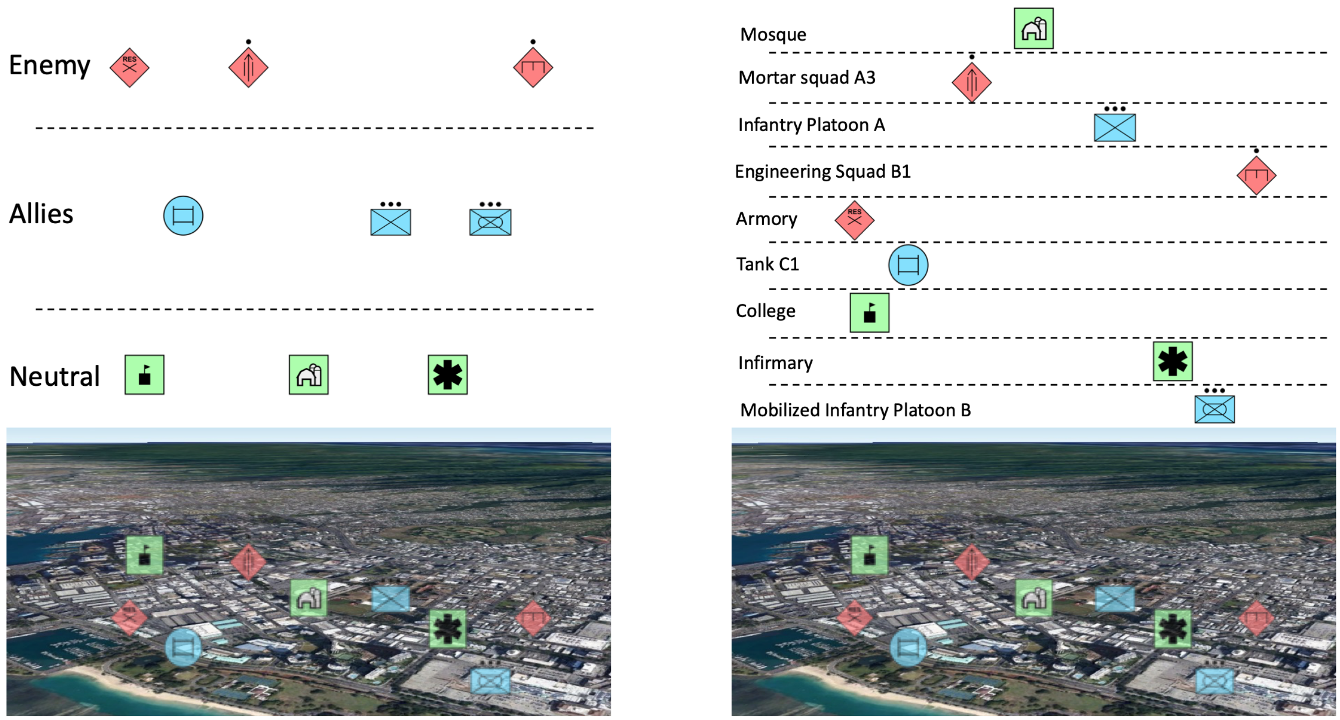

Figure 1: Left. An illustration of a layer-based filter, where entire layers can be turned on/off. Right. An information-item level based filter, where each single information item can be considered as a separate layer and handled individually.

illustrated in Figure 1). This can be achieved by an algorithm that dynamically and automatically filters information items on the C2 map by their importance and relevance to UAV operator's current mission. Automating this process, however, is challenging as it can inadvertently reduce operators' SA and performance, especially regarding items chosen not to be shown (Endsley and Kiris 1995).

Two guidelines led the development of the automatic and dynamic algorithm. First, working at an information item level requires advanced techniques. Those techniques rise from the artificial intelligence (AI) domain. Second, as the problem is both spatial and temporal, the solution should adopt perceptual concepts inspired by Gibson and Crooks' (1938) field of safe travel. The field of safe travel, was defined as the spatial field where it was safe to steer a car. It was dependent on driver state, vehicle and environment, and it was constantly changing as the vehicle moved and the context of driving changed. Gibson and Crooks referred to the field's intuitiveness as affordance. In Section 2 we use these guidelines for the development of the GiCoMAF (Gibsonian Command and Control Map AI Filter algorithm) but first, in Sections 1.1 and 1.2, we review Machine Learning (ML) models that were considered for the AI implementation and delineate the research goals.

1.1 Machine-Learning (ML) Models for C2 Systems

The military domain is seeking for techniques to incorporate AI into C2 systems and Machine Learning (ML) models have been used in UAV, GIS and C2 domains (e.g. Azak and Bayrak 2008; Bao 2016; Choi and Cha 2019; Dzieciuch et al. 2017; Noh and Jeong 2010; Rapaport 2015), but to our knowledge, not as the interface between the C2 map and the operators. For our research, we looked for models that can be used for classification (to classify *information item's importance*) or logistic regression (for defining the *field of relevance*, as detailed later). Table 1 describes four ML techniques, suitable for classification and regression and therefore applicable towards GiCoMAF development.

There are other ML models that can be used for the purpose of this study. For example, Long Short-Term Memory (LSTM), is a common deep learning technique for time series data, that can be used for multi-stream data as well (Behera, Keidel, and Debnath 2019; Bouaziz et al. 2017). Given that each information item in the C2 map can be perceived as a standalone time series, this model was considered for this study too. However, the exact number of parallel streams (i.e. the number of the information items) is constantly changing as more information items are added to the map. Therefore, the LSTM model was neglected.

1.2 Research Goals

Acknowledging UAV operators' C2 map needs, this study aims to introduce the GiCoMAF solution for distilling the information presented on the map to show mission relevant important information items, while minimizing or hiding less important or distracting items. The first goal is to develop an algorithm that automatically and dynamically fil-

Table 1: A short description of machine learning models suitable for classification and regression modeling.

Model	Description
Lasso Regression (LR)	In Generalized Linear Models (GLM) the relationship between the predicting variables and the independent variable is a linear function. It is mostly used to predict a single continuous numeric value, but variations allow it to be used for classification (multinomial linear regression; Greene 2012) or for predicting a probability (logistic regression; Walker and Duncan 1967). Lasso is a method of performing feature selection alongside GLM by introducing a coefficients size penalty as a constraint. The effect of the penalty on the regression is controlled by a user defined λ parameter (Tibshirani 1996; Tibshirani et al. 2012).
Neural Networks (NN)	NN is a composite of input, output, and middle layers (a.k.a. 'hidden layers') presuming to mimic the work of neurons in the human brain. Each layer consists of an undefined number of neurons, which sum the input received from the previous layer using an activation function f, and 'fire' the result to the next layer. NNs can handle non-linear problems (Kanevski et al. 2004).
Random Forest (RF)	Decision tree is a prediction for non-linear problems where each node of the graph represents a predicting variable. The returned value is the final node the decision led to (Quinlan 1986). The number of possible results is limited by the number of possible values of the label, and by the depth of the tree (maximum number of nodes a branch can have). Increasing the number of possible results can be done by increasing the tree's depth, but it may result in undesired patterns of overfitting. Random Forest overcomes this limitation by producing multiple trees, each one is trained on a randomly selected subset of the same dataset. The returned value is an average of the predictions of all trees for a regression type model, and the highest frequent value for the a classification type model (Breiman 2001).
XGBoost	Gradient boosting is an iterative ensemble of trees model, where in each iteration a decision tree is learned using the residuals of the previous iteration (Friedman 2002). XGBoost is an open source package implementing an efficient scalable tree boosting system, incorporating a regularized model to prevent overfitting (Chen and Guestrin 2016).

ters the information items on the C2 map. The *automatic* part of the algorithm, addresses the filtering at the information items' level. The *dynamic* part of the algorithm addresses the evolving environmental context of the area of operation. The second goal is to delineate the use of ML models in the construction of the algorithm by demonstrating how its construction can be achieved using ML models. This goal is attained using importance and relevance labelled data collected empirically from UAV operators. Their inputs enable the ML models to learn the operators' contextual information needs, and to showcase the algorithm's feasibility. This paper describes the process of developing the GiCoMAF algorithm using ML models. A third goal, not described in this paper, is to evaluate the update rate of the GiCoMAF empirically, exploring its effect on operators' mental workload, situation awareness and perception of the experience.

In the following Sections the definition of the GiCoMAF algorithm is first outlined. Then, the process of acquiring the ML models constructing the algorithm's tiers is detailed, including the data collection, manipulation, and models evaluation. Lastly, the discussion Section discusses how to combine all tiers into an operating filter algorithm.

2 Developing GiCoMAF

The GiCoMAF – Gibsonian Command and Control Map AI Filter algorithm, consists of two tiers, each answers a different research question. The integration of the tiers creates the filter rule, and incorporates the outcomes of these questions into the workflow of the operators. Tier I aims at the information item level, and answers the question *what is the perceived importance of each information item?* Tier II aims at the map as a whole, and answers the question *how can the operator's area of interest be modeled on the map?* Tier III aims at the dynamicity property of the algorithm, and answers the question *how often should the automatic filter be updated?*

The construction of the GiCoMAF is illustrated in Figure 2. The distinction between Tiers I and II is important. Tier I predicts each information item's importance. The scale is inclusive, i.e., it provides an indication of how important it is to <u>show</u> the information item on the map, and an indication of how important it is <u>not to show</u> the item. The incentive behind this logic is that some non-important information items may be harmless and operators will be oblivious to their presence, while others may be disturbing. Moreover, predicted information item's importance should not be handled in the same way throughout the map space, and this is where Tier II comes into play. Consider an information item with a neutral predicted importance, i.e., not important but not disturbing. While the item per se is not considered disturbing, possibly, if it is within the operators' area of interest, they may be more sensitive to disruptions, and the item can inadvertently cause clutter or quickly become disturbing. To avoid such cases, it may be better to filter out the item. Outside the area of interest, however, leaving a neutral item may be a good strategy, as its importance may rise as the mission evolves. Hence non-important items for the immediate context, may be valuable to foresee future evolvement of the situation and prepare for it. Therefore, Tier I of predicting the information items' importance is not enough, and the algorithm should model the operators' area of interest as derived in Tier II. The final decision rule, hence, is a combination of these two tiers. Tier I and II of the algorithm are based on ML models, thus, by learning examples from UAV operators, the algorithm predicts and executes an automatic filter for new operators and in new unknown scenarios. ML models require tagged examples to be learned from. Therefore, an experiment which emulated the work of UAV operator in operational scenarios was designed and conducted to collect tagged data from UAV operators (Zak, Parmet, and Oron-Gilad 2019a, 2019b). Tier III defines the rate at which the filter rule of Tiers I and II should be applied. At this stage of development Tier III cannot be based on ML models, and represent a pure cognitive issue. Since the scope of this study was to delineate the construction of the GiCoMAF algorithm using ML models, the process of studying the cognitive effect of various filter update rates and setting the optimal rate is due to future research.

2.1 Tier I – Information Item Importance

Tier I aims to model information items' importance as perceived by the operators. An information item's importance may vary based on environmental context, mission, and characteristics. Generally, operators want important information items to be shown on the map. If an information item is not perceived as important, its presence can be disturbing and then operators would prefer that it will not to be shown, or not disturbing and operators may be impartial to its presence on the map. Therefore, Tier I attempts to predict and classify information items' importance level, into a four tics scale, based on the context: Positive importance represents **important** (1) and **very important** (2) information items. Zero importance represent **non-important** information items, that operators have no preference to whether they should be shown or not. Negative importance represents information items that **disturb and distract** operators from their mission context.

The prediction is done using a ML model. The model, once trained, has the ability to deduce item importance from insightful environmental and mission related measures that describe the context. For example, the average distance of an information item from the UAV payload may describe its mission related context. The density of information items around an information item, for example, may reflect the environmental context. Higher density raises the probability of some operational event happening at that location. A ML model can find the relationship between an information item and the route, and then predict the importance of the information item; and it can determine that as the environment is denser, the probability of showing non-important items increases. These examples of relations between derived measures like distance and density and the predicted value are given for simplifying purposes, and the real relationship that emerge from the ML model may be more complicated. Furthermore, selecting the most suitable ML model (Table 1) was determined empirically as detailed in Section 3.

2.2 Tier II – Operator's Field of Relevance

Tier II of the algorithm addresses the areas of interest for the operators in the environment. It is essentially a spatial problem that can be represented using geospatial measures on the map (e.g. polygons, heatmaps, etc.). The 'area of interest' is not necessarily a singular area, and can be phrased as 'areas of interest', where each area has a different level of interest. For example, Figure 3 illustrates multilevel areas of interest, where in the center of the polygons occurs the mission, and therefore the smallest area around this focus has the highest interest. The surroundings do withhold interest to the operator since they may affect the mission's center. Their relevance decreases as they get farther than the mission's center. The 'area of interest' is modeled using the concept of *field of relevance*, an adaptation of Gibson and Crooks (1938). The field of relevance depicts operators' areas of interest based on the environment, mission, operators' behavior, etc. Moreover, it corresponds with the affordance property as the field of relevance highlights areas that are intuitively more focused upon by the operators. Due to the probabilistic characteristic of ML models, it was decided to adopt the quality map approach of Morse, Engh, and Goodrich 2010, and to model the field of relevance as a pseudo-Gaussian heatmap, where each spatial element on the map gets a value between 0 (no relevance) and 1 (high relevance) corresponding to the probability of that spatial point to be in the operator's area of interest.

Similar to Tier I, prediction in this tier is done using a ML model, deducing from insightful environmental and mission related measures that describe the context. In this tier those measures are in respect to a spatial element (e.g., a square of 10 meters^2), without relating to any particular information item within the area of interest. For example, a mission related measure can be the average time a spatial element is in the UAV payload's field of view, assuming higher average time may indicate higher relevance of that element. An environmental measure can be the density of information items around a certain element. Assuming higher density around a spatial element indicates upon the probability of some operational event happening at that location, and in turn higher relevance. Similar to Tier I, the examples of relations are for simplifying purposes, and the exact ML model (Table 1) was determined through an empirical process detailed in Section 3.

3 Constructing the GiCoMAF

In the construction of the filter, an experiment emulating the work of UAV operators in the military domain was executed. Participants, professional military UAV operators, were asked to perform a mission of supporting a ground battalion in urban battlefield scenarios. The data collected using the feedbacks they provided during and after the experiment was used to construct the ML models of Tiers I and II. The process is illustrated in Figure 2. The research was approved by the Institutional Review Board at Ben Gurion University. Informed consent was obtained from each participant.

3.1 Data Tagging Experiment

Data for Tiers I and II were collected in a set of experimental runs, detailed in Zak, Parmet, and Oron-Gilad 2019a and 2019b. The experiment aimed to emulate the work of UAV operator in the military domain. Using a designated system developed for this task (UCES – UAV Command and Control Experiment System), a battlefield scenario was developed by subject matter experts with a UAV mission to

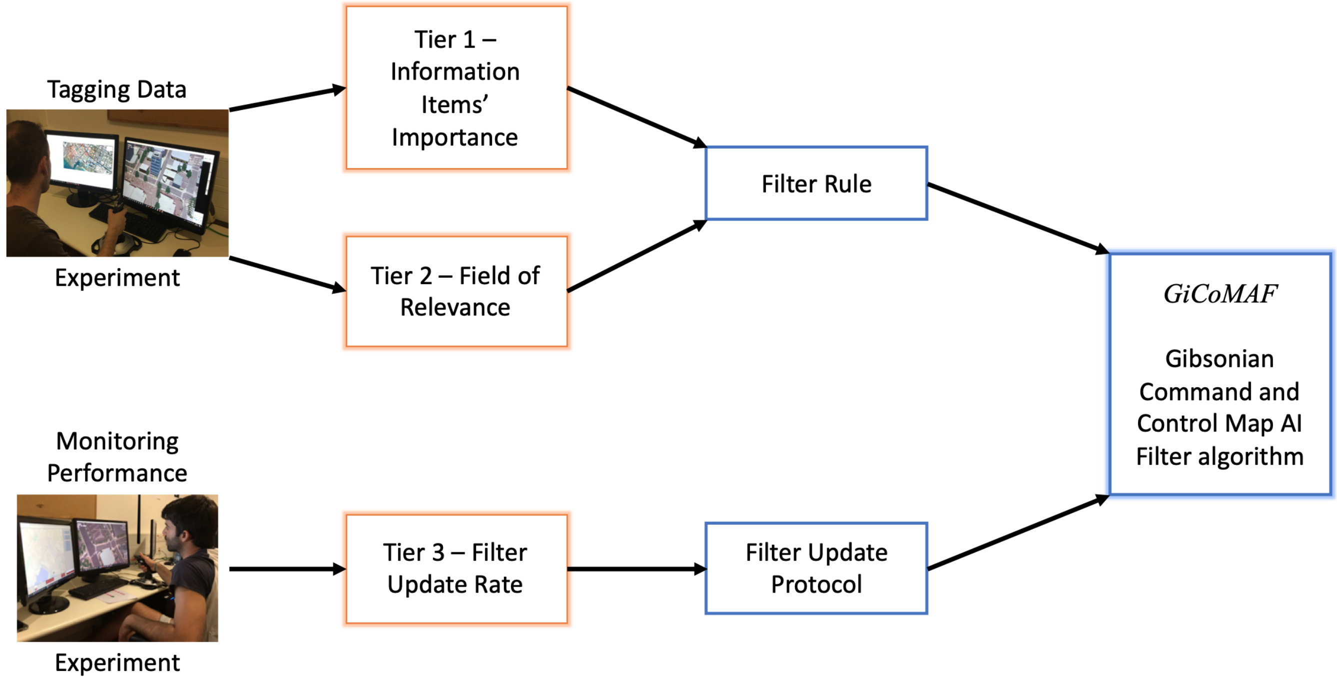

Figure 2: The construction process of the GiCoMAF. Tier I and II provide the decision rule for what should be presented on the C2 map. Tier III is related to the update rate of the filter and how it affects operators' mission performance, workload and experience.

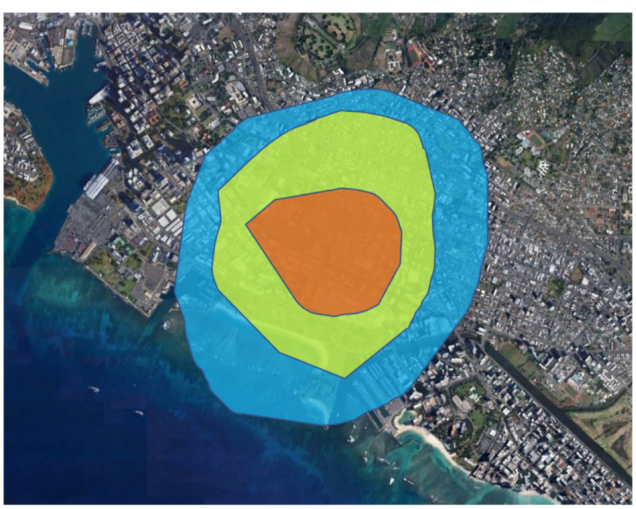

Figure 3: An illustration of multilevel areas of interest, as derived by operators in an operational context. Each polygon represents a level of interest, increasing from the outside-in.

assist a ground battalion in conquering an urban neighborhood. The 'five paragraph order', a common military standard of writing a fight plan of the mission, was outlined on the C2 map (Figure 4) and programmed in the VR-Forces simulation engine. The mission was 12 minutes long, representing a sequence of events that in real life settings may take several hours. Thirteen professional military UAV operators performed a reconnaissance mission as if they were acting in a real-world battlefield. The UCES allowed them to control the vehicle's payload, observe the battlefield from an aerial point of view, and get real-time information from a C2 map. Occasionally at specific points, the scenario paused, and a scoring session had started. They were asked to label two types of information using the UCES map: tagging individual information items' importance; and drawing their current contextual area of interest as polygons on the map. There were 88 scoring sessions in total, an average of 6.5 sessions for each experimental run. The data collected from the runs was put together into two datasets. The dataset containing the information item's importance tags was the input for Tier I, and the dataset containing the area of interest polygons was the input for Tier II. Then, ML models for the two tiers were developed. The processes and outcomes are detailed in Sections 3.2 and 3.3.

3.2 Tier I ML Model – Information Items' Importance

The data collected in the experiment was divided into two tables. The first table was an event diary, where each row

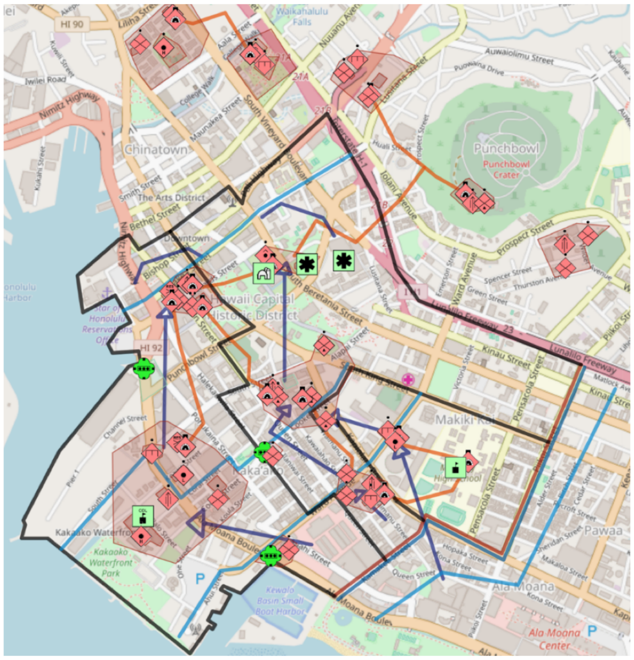

Figure 4: A mockup of the C2 map, including the fight plan for the simulation's scenario (in dark blue). The symbols are standard NATO symbols, where light blue represents allied forces, red represents enemy forces, and green represents neutral entities.

represents one event in one experimental run (e.g., UAV movement, forces movement, cross-fire event, etc.). This table contained approximately 90,000 rows. The second table is a tagging table, where each row represents an information item in a scoring session in an experimental run and its perceived importance as was reported by the participant. This table contained 1,203 rows, where each information item has inherent attributes like its location, type, the time it was last modified, etc. However, item attributes are insufficient for introducing mission context to the ML models. Therefore, the two tables had to be joined into one training dataset. For that, the event diary had to be manipulated into insightful variables. Thirty-five derived measures were calculated from the event diary to describe the environment and the mission. The objective of the ML model was to classify information items' importance. Misclassifying a non-important information item as important does not have the same implication as misclassifying an important information item as disturbing. Therefore, the error function could not be merely classification accuracy, and weights were given as penalty for misclassifications based on the sensitivity and specificity of the misclassification.

All four techniques described in Table 1 were tested in a classification configuration. Model technique parameters were optimized using the k-fold cross validation technique, where $k = 10$. During this process, 22,945 models were built in total for all four techniques. Then, data were divided by participant; 10 randomly chosen participants as the training set and the 3 remaining as the validation set. The final model for each technique was built using the derived optimal parameters set. Figure 5 illustrates the results for two types of errors: (a) weighted error, the average k-fold result for the optimal parameters set, the training set, and the validation set; and (b) sign error (percent of misclassifications of important/very important as disturbing, and vice versa), for all three sets. From Figure 5 it is evident that RF shows patterns of overfitting, and multinomial LR and NN perform better than XGBoost in terms of weighted error. However, in terms of sign error XGBoost outperforms the others. Figure 6 delves into the differences among the models by illustrating a multiclass receiver operating characteristic (ROC) graphs (Hand and Till 2001) on the validation set. In this figure, each line represents the ROC of the comparison of two possible values, and the area under the curve (AUC) is the average of all AUCs. It can explain the differences between weighted error and sign error patterns of the NN and LR. As seen in the plot, NN and LR have almost no separation between the predicted values, probably because misclassifying very important information items as disturbing had the highest penalty. Thus, both models tend to classify all information items as 'very important'. Moreover, according to Figure 6, RF had slightly better AUC, although Figure 5 suggests it was overfitting. A third analysis to evaluate the models' performance was to look at how each model was certain in the prediction of the validation set. In the prediction process, each model provides a probability for each level, where the predicted value is set by choosing the level with the highest probability. It is expected that a more 'decisive' model would provide relatively high probability for the predicted value and relatively low probabilities for the other values. The entropy for that prediction, defined by the formula $-\sum_i p_i log_a p_i$ (where p_i is the probability of predicting a given value i, and a is the number of possible values), would then be close to zero. The result of the entropy analysis as given in Table 2, show the average entropy for each model and predicted value. It is evident from Table 2 that LR assigns only the value '2', and NN only the values '1' and '2'. All four models have relatively high entropy scores, sometimes close to one, indicating that the models' classification decisions, based on the levels' probabilities, are hung on the fluctuation of an ϵ. I.e., and all four models are not very decisive. The only exception is the RF model, which had a relatively lower entropy, although still closer to one than zero.

Table 2: Entropy analysis for each model.

Model	Predicted Value			
	Negative	0	1	2
LR	NA	NA	NA	0.921
NN	NA	NA	0.891	0.917
RF	0.716	0.875	0.739	0.661
XGBoost	0.999	0.999	0.999	1

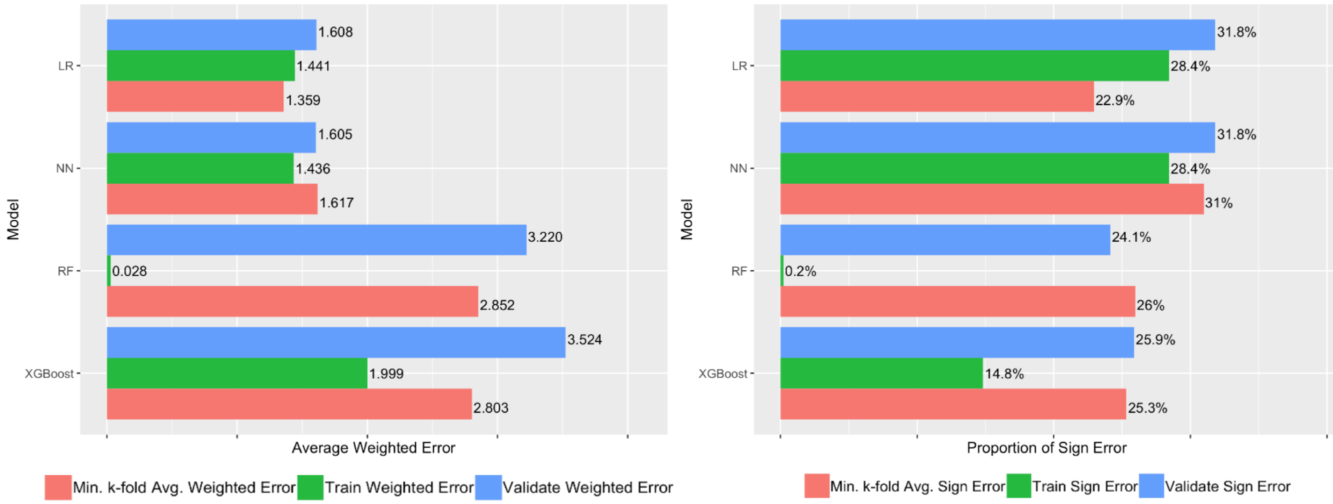

Figure 5: ML model performance result for Tier I. Left. The average weighted error of the k-fold result of the optimal parameters set (red), the weighted error of the training (green) and the validation (blue) sets. Right. The corresponding sign error of the k-fold and the sets. One can see that LR and NN models provide less weighted error in comparison to RF and XGBoost, but more sign error.

3.3 Tier II ML Model – Field of Relevance

The objective of the model was to define the field of relevance by predicting the probability of each point on the map to be in the area of interest. To facilitate the model construction, the C2 map was divided into a grid of cells, where each cell represents an area of 25x25 meters on the ground, 9,016 cells in total. Each row in the dataset represents one cell in a scoring session of an experimental run, with the predicted label of (1) if the cell was in the drawn area of interest in that scoring session, and (0) otherwise. The derived measures were calculated with respect to each cell (e.g. the average distance of a cell from the UAV route, etc.).

Similar to Tier I, the data was divided into two tables; an event diary and a table with the reported feedback from the participants. In this tier, each row in the reported feedback table represents coordinates of the area of interest's polygon of a scoring session in an experimental run. Each scoring session consisted of one polygon. Subtracting two cases where the participant forgot to draw an area of interest, this table had 86 rows. Also, the events diary was manipulated into 20 derived measures.

The objective of the model was to define the field of relevance by predicting the probability of each point on the map to be in the area of interest. To facilitate the model construction, the map was divided into a grid of cells, where each cell represents an area of 25x25 meters on the ground, 9,016 cells in total. Thus, each row in the dataset represents a cell in a scoring session of an experimental run, with the predicted label of (1) if the cell was in the drawn area of interest in that scoring session, and (0) otherwise. The derived measures were calculated with respect to each cell (e.g. the average distance of a cell from the UAV route, etc.).

The same four ML techniques that were used in Tier I were used in the regression configuration, but the predicted value was logistic (a number in [0,1]). Model performance was measured using root mean square error (RMSE). Due to the large scale of data and computing resource limitations, 408 models were built in total in four k-fold cross validation processes, where $k = 3$. Figure 7 illustrates the average k-fold RMSE results for the optimal parameters set, and the RMSE of the training and validation sets. While logistic LR and NN have better results on the validation set in terms of RMSE, both XGBoost and RF perform better on both the training set and k-fold results. Figure 8 delves into the results and illustrates a 'rotated confusion matrix'. The x-axis represents the predicted value, binned in intervals of 0.01. The y-axis represents the average original value of the rows corresponding to the bins of the x-axis. The size of the points represents the relative number of records in each bin. A perfect model would provide points aligning with the diagonal of the plot. According to Figure 8, both LR and NN collapse and do not provide good differentiation. RF provides something close to binary results, which may conflict with the field of relevance concept. Therefore, although not perfectly aligning with the diagonal, XGBoost performs better than the other three models. An illustration of XGBoost's performance relative to the runner-up RF model, in two scoring session of two different participants is given in Figure 9. Figure 9 demonstrates how each operator marked the field of relevance, at the same stage showing that the structure of the field of relevance depends on the environment, the mission, and operators' individual characteristics and preferences. The XGBoost model seems to handle these characteristics well.

4 Discussion

Operators of UAVs work in uncertain and dynamic environments. Their main focus is on managing their vehicle's

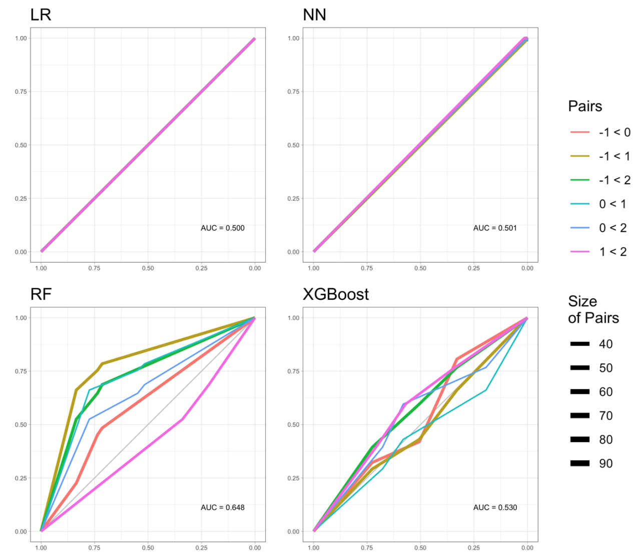

Figure 6: Multiclass ROC curves for the four ML models on the validation set, with the average AUC for each model. X-axis – False Positive Rate (1 – Specificity). Y-axis – True Positive Rate (Sensitivity). The width of each line represents the number of observations used for the development of the line in the curve.

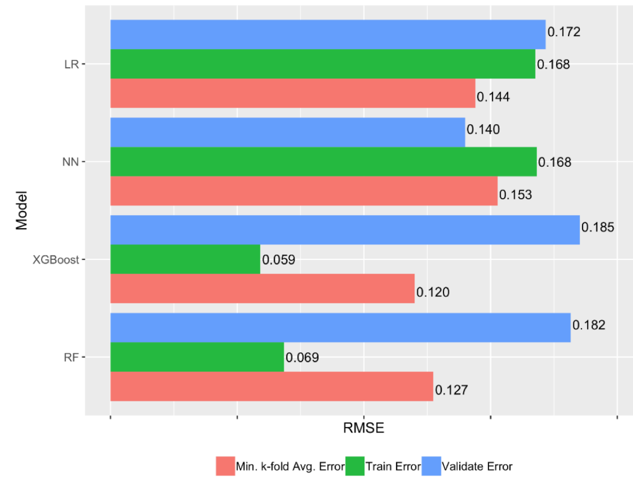

Figure 7: ML model performance result for Tier II. The average RMSE of the k-fold result of the optimal parameters set, and the RMSE of the training and the validation sets. LR and NN have better results on the validation set, both XGBoost and RF perform better on the training set and k-fold results.

payload (e.g., camera video-feed) while they are required to maintain orientation and awareness to their current location and its surroundings, and plan ahead. In order to develop the essential SA, they continuously and constantly interact with a C2 map (see Back et al. (2019) for a Cognitive Work Analysis). The map, is often overloaded with data that is irrelevant to their mission, and like the operational environment things change rapidly. The existing layer-based map filter mechanism causes a tradeoff dilemma for the operator –showing an entire layer with its relevant and irrelevant information, or hiding an entire layer and losing important information. Furthermore, as noted in Back et al. the existing filtering mechanisms require operators to manually interact with the interface, highlight or hide layers of information at the busiest times of their mission. Thus, currently, obtaining information from C2 maps pose high demands on operators, and therefore, they often cope with impaired SA. This study proposes a solution to dynamically and automatically adjust the C2 map display to operators' needs by introducing an AI-based dynamic and automatic algorithm that filters the information on the C2 map at the individual items level (as opposed to layers), as described in section 2, and laying the foundation for the algorithm by delineating the use of ML models in its construction as detailed in Figure 2.

The data tagging experiment was designed to collect labelled data towards the construction of the ML models of Tiers I and II; predicting information items' importance and predicting the field of relevance, respectively. Using the collected data, four different models were optimized and developed for each tier. In Tier I both XGBoost and RF had good results on the validation set, with RF on the upper hand. However, since the RF had extremely low training error, for both error types, it may be overfitted. Further study on the effects of each model on UAV operators' mission performance is needed to examine if the effect of overfitting is negligible. Therefore, to be cautious, the XGBoost model was chosen for further development. The XGBoost model, however, had an accuracy sign error of 25.9%, and its decisions were hanging on the thread of an ϵ (Table 2). This magnitude of the error is reasonable when modeling human preferences and performance (as seen in previous studies, e.g. Agichtein et al. 2016; Guimerà et al. 2012; Liu, Bian, and Agichtein 2008), especially when having a limited number of observations. The XGBoost results of Tier II were visually evaluated and provided the best and most accurate pseudo-Gaussian heatmap for the field of relevance. as illustrated in Figures 8 and 9. It should be noted that accuracy can be further improved by testing other ML models and parameters. Indeed, due to resource limitations, not all suitable models could be tested. Future study can attempt to improve the results of this study by applying additional ML models and techniques.

The two tiers delineate the use of ML models in the GiCoMAF algorithm . Yet, is not clear cut how to combine Tiers I and II into the GiCoMAF algorithm. Several options for how the tier combination protocols are being discussed in the following section.

4.1 Putting It All Together – Combining Tiers I and II

The exact combination of Tiers I and II into the GiCoMAF algorithm may depend on the mission, the environment, and even the organization that the operators are part of. This section suggests three possible combination approaches; however, the final setup is not limited to these approaches and

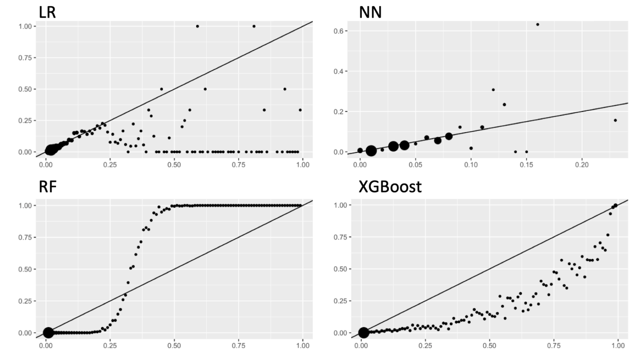

Figure 8: Predicted values binned into intervals of 0.01 (X-axis), and the average original values (Y-axis) of corresponding records. A perfectly fit model should be aligned with the diagonal line. The XGBoost model seems to be the better predictor of the four techniques.

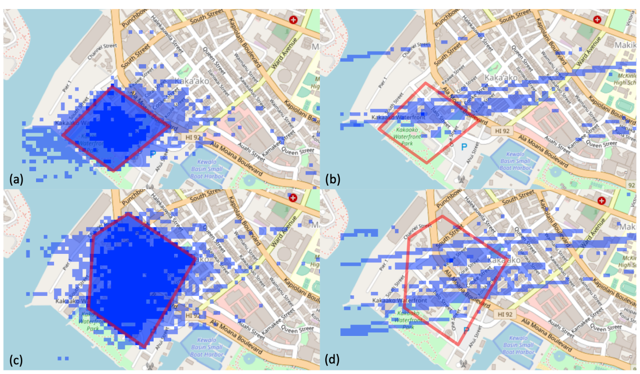

Figure 9: The field of relevance simulated in one scoring session in the validation set. The predicted field of relevance is illustrated in blue and the operator's reported area of interest is shown in red. Subfigures (a)-(b) and (c)-(d) represent two different participants, in the same part of the scenario. (a)-(c) are the predictions of the XGBoost model, (b)-(d) are of RF model.

each organization/user/system case may dictate the development of a tailored solution. These approaches provide a show/no-show decision rule based on relevance and importance thresholds. The optimal approach used in the algorithm, as well as its thresholds, should be continuously evaluated through a process of experiments.

Positive Correlation – This approach assumes that there should be a positive correlation between the field of relevance and the information density. Thus, the more relevant a spatial element is, the more information should be presented around it. To avoid distraction, the farther away from the spatial element, the less information should be shown. Figure 10a illustrates this approach where there is a negative relation between the field of relevance and the information items' importance. In order to use this approach, an adaptive show/no-show threshold should be set based on information items' importance, where the threshold is rather low at points with high relevance (implying more information items would be shown), and rather high at points with low relevance (only very important information items would be shown).

Negative Correlation – In contrast to the *positive correlation* approach, this approach assumes that there should be a negative correlation between the field of relevance and information density. I.e., the more relevant a spatial element is, the more attention operators direct to it, and therefore fewer information items should be shown to avoid disruption. Thus, as illustrated in Figure 10b, there should be a negative relation between the field of relevance and information items' importance, in similar but opposite approach to the relation described in the *positive correlation*.

Binary Relevance Decision – This approach assumes that only important information items should be shown on the map, and only in areas which are currently most relevant to the operators. Therefore, there are only two constant thresholds in the algorithm – one to decide the minimum importance an information item should have in order to be shown, and one to decide the minimum relative relevance a spatial point should have in order to set the importance threshold into motion. Figure 10c illustrates the approach.

For a further study, we propose to examine these three approaches and choose the fittest approach for the context that was addressed in this paper (i.e., the same mission type and profile of participants). For that a new multistage scenario similar in nature to the experimental scenarios detailed in this paper is required. Using the ML models developed in Tiers I and II, each scenario can be run in one of the three combination approaches, as well as running with no filter at all. In the first experiment, participants will be introduced to all of the approaches and different thresholds setups would be tested. Then, using the optimal threshold for each approach, new operators would participate in a two-stages crossover experiment with eight possible treatments – the three combination approaches and no filter at all.

5 Conclusions

This study aimed to introduce a solution to the information overload of C2 maps used by UAV operators. By solving operators' need for distilled information at the right time and in the right place, it is expected that operators will benefit more from the C2 map at lower efforts, and overall mission performance will improve. The study had met its goals. First, a solution was achieved by introducing the three-tier Gibsonian Command and Control Map AI Filter algorithm (GiCoMAF). Then, ML models were developed and evaluated using tagged data collected in an experiment. The results of the experiment are encouraging. The ML models that emerged from the first experiment were satisfying, and indicated high accuracy and usability of the algorithm, a step towards a solution to the information overload problem, and high workload of UAV operators. The combination of the tiers into a single algorithm is yet to be fully defined, and should be determined by an additional set of experiments we recommend performing in future studies.

The contribution of this study lays in various aspects. First, this study targets a long-neglected field of improving

Figure 10: An illustration of three possible algorithm combination approaches for Tier I and II. The heatmap represents the field of relevance, dimmed information items represent items that would not be shown to the operator.

the use of C2 maps by operators. Second, by using ML models in the construction of the GiCoMAF algorithm, this study utilizes AI concepts and techniques to lay the foundation of an algorithm that is targeted for improving human performance and human cognitive aspects of workload and SA. And third, successful construction of the algorithm can improve mission performance and enhance the cognitive abilities of operators to perform spatial-temporal tasks beyond the UAV domain. Algorithms of this form can be further implemented in any domains where an overloaded spatial and temporal information has to be filtered, e.g., emergency dispatching systems, information guided surgeries, search and rescue missions, air traffic control, etc.

6 Acknowledgments

This research is partially funded by the ”Negev” scholarship, and by the George Shrut Chair in Human Performance Management at Ben-Gurion University of the Negev. Corresponding author – Yuval Zak

References

Adams, J. A. 2015. Cognitive Task Analysis for Unmanned Aerial System Design. In *Handbook of Unmanned Aerial Vehicles*. Dordrecht: Springer Netherlands. 2425–2441.

Agichtein, E.; Brill, E.; Dumais, S.; and Ragno, R. 2016. Learning User Interaction Models for Predicting Web Search Result Preferences. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’06*. Seattle, Washington, USA: ACM.

Azak, M., and Bayrak, A. E. 2008. A new approach for Threat Evaluation and Weapon Assignment problem, hybrid learning with multi-agent coordination. In *2008 23rd International Symposium on Computer and Information Sciences, ISCIS 2008*.

Back, Y.; Zak, Y.; Parmet, Y.; and Oron-Gilad, T. 2019. Using Cognitive Work Analysis to Understand UAS Operators’ Map Display Needs. *Submitted to Requirements Engineering, April 2019*.

Bao, T. 2016. Swimming In Sensors , Drowning In Data — Big Data Analytics For Military Intelligence. *Pointer, Journal of the Singapore Armed Forces* 42(1):51–65.

Behera, A.; Keidel, A.; and Debnath, B. 2019. *Context-driven Multi-stream LSTM (M-LSTM) for Recognizing Fine-Grained Activity of Drivers*, volume 11269 LNCS. Springer International Publishing.

Bouaziz, M.; Morchid, M.; Dufour, R.; Linarès, G.; and De Mori, R. 2017. Parallel Long Short-Term Memory for multi-stream classification. *2016 IEEE Workshop on Spoken Language Technology, SLT 2016 - Proceedings* 218–223.

Breiman, L. 2001. Random Forests. *Machine Learning* 45(1):5–32.

Calhoun, G. L.; Draper, M. H.; Abernathy, M. F.; Patzek, M.; and Delgado, F. 2005. Synthetic vision system for improving unmanned aerial vehicle operator. In *Proc. SPIE 5802, Enhanced and Synthetic Vision 2005*, volume 5802, 219–230.

Chen, T., and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. In *KDD ’16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

Choi, S. Y., and Cha, D. 2019. Unmanned aerial vehicles using machine learning for autonomous flight; state-of-the-art. *Advanced Robotics* 33(6):265–277.

Dzieciuch, I.; Reeder, J.; Gutzwiller, R.; Gustafson, E.; Coronado, B.; Martinez, L.; Croft, B.; and Lange, D. S. 2017. Amplifying Human Ability through Autonomics and Machine Learning in IMPACT. In *Proc. SPIE 10194, Micro- and Nanotechnology Sensors, Systems, and Applications IX*.

Endsley, M. R., and Kiris, E. O. 1995. The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors* 37(2):381–394.

Endsley, M. R. 1988. Situation awareness global assessment technique (SAGAT). *Aerospace and Electronics Conference, 1988. NAECON 1988., Proceedings of the IEEE 1988 National* 789–795.

Endsley, M. R. 2000. Theoretical Underpinnings of Situation Awareness: A Critical Review. In *Situation Awareness Analysis and Measurement*. Lawrence Erlbaum Associates. 3–32.

Everaerts, J. . 2008. The Use of Unmanned Aerial Vehicles (UAVs) for Remote Sensing and Mapping. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVII(Part B1):1187–1192.

Filkins, D. 2010. Operators of Drones Are Faulted in Afghan Deaths.

Friedman, J. H. 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis* 38(4):367–378.

Gibson, J. J., and Crooks, L. E. 1938. A Theoretical Field-Analysis of Automobile-Driving. *The American journal of psychology* 51(3):453–471.

Greene, W. H. 2012. *Econometric Analysis (Seventh ed.)*. Boston: Pearson Education.

Guimerà, R.; Llorente, A.; Moro, E.; and Sales-Pardo, M. 2012. Predicting Human Preferences Using the Block Structure of Complex Social Networks. *PLoS ONE* 7(9):3–9.

Hand, D. J., and Till, R. J. 2001. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning* 45(2):171–186.

Izzetoglu, K.; Ayaz, H.; Hing, J. T.; Shewokis, P. A.; Bunce, S. C.; Oh, P. Y.; and Onaral, B. 2015. Uav Operators Workload Assessment by Optical Brain Imaging Technology (fNIR). In *Handbook of Unmanned Aerial Vehicles*. Dordrecht: Springer Netherlands. 2475–2500.

Kanevski, M.; Parkin, R.; Pozdnukhov, A.; Timonin, V.; Maignan, M.; Demyanov, V.; and Canu, S. 2004. Environmental data mining and modeling based on machine learning algorithms and geostatistics. *Environmental Modelling and Software* 19(9):845–855.

Liu, Y.; Bian, J.; and Agichtein, E. 2008. Predicting Information Seeker Satisfaction in Community Question Answering. *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '08* 483.

Marusich, L. R.; Bakdash, J. Z.; Onal, E.; Yu, M. S.; Schaffer, J.; ODonovan, J.; Ho llerer, T.; Buchler, N.; and Gonzalez, C. 2016. Effects of Information Availability on Command-and-Control Decision Making: Performance, Trust, and Situation Awareness. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 58(2):301–321.

Morse, B. S.; Engh, C.; and Goodrich, M. A. 2010. UAV video coverage quality maps and prioritized indexing for wilderness search and rescue. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 227–234. Osaka: IEEE.

Noh, S., and Jeong, U. 2010. Intelligent Command and Control Agent in Electronic Warfare Settings. *International Journal of intelligent Systems* 25:514–528.

Quinlan, J. R. 1986. Induction of Decision Trees. *Machine Learning* 1(1):81–106.

Rapaport, A. 2015. "Quite a few Terrorists lost their lives owing to Big Data".

Sandom, C. 2000. Operator Situational Awareness and System Safety. In *IEE One Day Seminar on Systems Dependency on Humans*, volume 2000, 5–5. IEE.

Shanker, T., and Richtel, M. 2011. In New Military, Data Overload Can Be Deadly.

Tibshirani, R.; Bien, J.; Friedman, J.; Hastie, T.; Simon, N.; Taylor, J.; and Tibshirani, R. J. 2012. Strong Rules for Discarding Predictors in Lasso-type Problems. *Journal of the Royal Statistical Society. Series B (Methodological)* 74(2):245–266.

Tibshirani, R. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1):267–288.

Walker, S. H., and Duncan, D. B. 1967. Estimation of the probability of an event as a function of several variables independent. *Biometrika* 54(1 and 2):167–179.

Zak, Y.; Oron-Gilad, T.; and Parmet, Y. 2018. Operator Workload Reduced in Unmanned Aerial Vehicles : Making Command and Control (C2) Maps More Useful. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*.

Zak, Y.; Parmet, Y.; and Oron-Gilad, T. 2019a. Making Command and Control Maps More Useful for Operators of Unmanned Aerial Systems : A Preliminary Model and Empirical Approach. In *HSI2019*.

Zak, Y.; Parmet, Y.; and Oron-Gilad, T. 2019b. Towards the Development of a Display Filter Algorithm for Command and Control (C2) Maps for Operators of Unmanned Aerial Systems. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19. Glasgow, UK: ACM.