Early Information Parameter-Set Analysis for Satellite Close Approaches using Machine Learning

Brianna I. Robertson* and Alinda Mashiku, PhD†
NASA Goddard Space Flight Center, Greenbelt, MD, 20771

Understanding orbital mechanics is essential in space flight and navigation applications, and leveraging modern force models for flight path projection remains an important aspect in space mission design and operation. However, force models do not capture all the dynamics or perturbations in the space environment and thus are subject to errors in predicting the state vectors. The further out the predicted miss distance between spacecraft is from the time of closest approach (TCA), the larger the propagated errors in the predicted miss distance at TCA is. The dependency on these force models for spacecraft flight state prediction calls for a more reliable method that can quantify, or even reduce, these propagated errors. With recent advances in the field artificial intelligence, specifically in machine and deep learning algorithms, a model that implements these approaches can improve on the modern force model approach. The goal for this work is to provide an early-information decision-making threshold, in order to prioritize risk assessment implementation, given the ongoing increase of space objects. In analyzing the relationship of several parameters from conjunction data messages (CDMs) and solar information, early information becomes viable in miss distance prediction with unsupervised learning techniques, which learn the parameters that are linked together with miss distance and probability of collision (Pc) variables. Another approach implemented for identifying relationships within CDMs is supervised learning, in which a shallow neural network binary classifier learns to distinguish events with Pc values $> 10^{-8}$. These parameters detected in the unsupervised process are then applied to a regression neural network, which predicts the miss distance at TCA for a specific event within a given uncertainty bound. For the regression neural network, a Long Short Term Memory (LSTM) neural network is implemented, which yields memory about each time step in an event. Using an LSTM network, the model learns to predict miss distance within 0.2km of the value measured at TCA. Although there is a limited amount of "close miss" data to train a network, the network learns to associate parameters, like large energy dissipation rates with the secondary object, with an elevated Pc.

Keywords - Orbital Mechanics, Collision Avoidance, Machine Learning, Unsupervised Learning, Long Short Term Memory

I. Nomenclature

CDM	=	Conjunction Data Message
TCA	=	Time of Closest Approach
Pc	=	Probability of Collision
LSTM	=	Long Short Term Memory
LEO	=	Low Earth Orbit
CARA	=	Conjuction Assessment and Collision Avoidance
$md\sigma$	=	miss distance standard deviation measured across event
RNN	=	Recurrent Neural Network

*Senior, Louisiana State University, Computer Engineering & Physics
†NASA CARA Deputy Manager, Code 595

II. Introduction and Background

For decades, the "big sky theory," the idea that the amount of available space in Earth's orbit was too large to seriously consider the possibility of collision, dominated the space industry, negating the need for careful risk assessment and mitigation [1]. However, the "big sky theory" no longer applies in the same capacity. As the number of objects in orbit increases, as does the need for early risk assessment and decision. Ongoing projects, such as SpaceX's Starlink initiative, adds thousands of more objects to the current sky, increasing the need for state-of-the-art risk mitigation practices [2]. Furthermore, recent collisions, like the Iridium-33/Cosmos 2251 in 2009, exponentially increased the space debris that requires both tracking and collision assessment [3]. Over 500,000 pieces of debris orbit the Earth, and this number is expected to continue to grow as collision risks increase [4]. To combat potential collisions, NASA developed the Conjunction Assessment and Collision Avoidance (CARA) team , which is currently responsible for defining the metrics for safe and unsafe approaches. Implementing orbital determination models, CARA works with information as early as seven days before a close approach, but the decision to implement a maneuver essential to preventing a collision still only occurs within a couple days of the closest distance between the two objects. Moreover, the predicted radial distance between two observed objects is subject to large error, especially in events with known perturbations and non-Gaussian covariances; therefore, maneuvers, implemented changes in the orbit of a satellite to prevent collisions, are difficult to implement sooner [5, 6].

A. Current Methods for Probability of Collision (Pc) Measuring

There are several different methods for measuring the probability of collison (Pc) of an event. Conjunction Data Messages (CDMs) contain over 200 points of information about a primary and a secondary object's relationship, including state vectors, probability of collision (Pc), and time of closest approach (TCA). These are collected three times daily, with the earliest data collection for an event about seven days before TCA. Using standard orbital determination models, the trajectory of the objects are calculated and screened for near approaches. This screening process is standardized through the use of Monte-Carlo simulations, as discussed in [5], which intakes the positions, velocities, and covariances of these parameters to construct the Pc value. These current models lack confidence in predicting the Pc and miss distance of an event that is several days before TCA, mainly due to perturbations and other unaccounted reasons. As the number of objects in orbit increases, the importance of early collision detection and avoidance amplifies, for early information may prevent risk mitigation from taking place too late in the event timeline.

B. Covariance Realism

One of the problems with predicting miss distance is covariance realism, which is when the measured covariance of the state vectors is much larger or smaller than the actual error volume [5]. Inaccurate covariances can lead to misguided risk margins for events, leading to the prioritization of a non-close approach event over a legitimate risk. Moreover, the assumption of Gaussian-error ellipsoids can be dangerous, as nonlinear dynamics effect spacecraft trajectory [6]. Although current models for miss distance prediction error bound the prediction, the error bars may not encapsulate the entire risk. Although current methods have been effective in risk mitigation for objects in orbit, a more accurate model that predicts at an earlier stage of an event timeline accurately is necessary for proper conjunction assessment and risk analysis.

C. Machine Learning and Risk Assessment

The emerging success of machine learning to regression and classification applications continues to impact the scientific community. Applying machine learning to early collision risk assessment can enhance the general understanding of what parameters are associated with poor miss distance prediction. Moreover, identifying these trends early in flight path trajectory enables early-risk mitigation, especially in the case that a maneuver is implemented. Increased collision risk parameter association and early-information for decision making benefits from both unsupervised learning and supervised learning. Unsupervised learning algorithms interpret information about the data set without having a target to achieve, while supervised learning models are given both data and a target. For a supervised neural network model learns to associate the given data with the target, minimizing loss, which is the difference between model output and target value. Unsupervised learning, on the other hand, has no target; therefore, many unsupervised learning algorithms work using cluster identification.

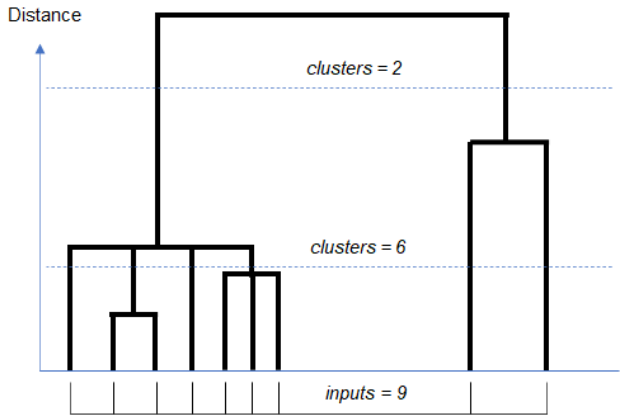


Fig. 1 Hierarchical clustering results displayed in a dendrogram. In this example, nine inputs are clustered. Horizontal slicing for the number of clusters = 2, 6 cases are indicated.

D. Unsupervised Learning

In this work, unsupervised learning is used for feature selection, particularly with clustering and competitive network methods. Clustering algorithms, which span across a data set to identify regions where data is clustered together. The clustering method used in this work to associate CDMs is agglomerative hierarchical clustering, which has success in high-dimensionality reduction and feature extraction in other applications[7, 8]. Due to the number of dimensions assessed for probability of collision, hierarchical clustering was selected to identify clusters to reduce the number of dimensions used for risk mitigation. Hierarchical clustering works bottom-up, where each data point is associated with its neighbor with the closest norm distance [9]. The results of this operation are displayed in a dendrogram, which is a tree-like structure as seen in Fig. 1. At each level of a dendrogram, a horizontal slice reveals the number of clusters and points in a given cluster. For this work, hierarchical clustering identifies parameters with the probability of collision and miss distance metrics.

Another unsupervised learning method utilized in this work is the competitive networks, which initializes a number of weight vectors that train on the data as it is inputted to the network. This iterative process takes an input data, such as a CDM, and identifies the weight vector that is closest to that inputted data. That weight vector updates via the Kohonen rule 1, [10].

$$ {}_i\mathbf{w}(q) = {}_i\mathbf{w}(q-1) + \alpha(\mathbf{p}(q) - {}_i\mathbf{w}(q-1)), \quad \text{for } i \in X(q) \tag{1} $$

where $\mathbf{w}$ represents the weight vector, $\mathbf{p}$ represents the input, α is the learning rate, i is the weight indice, and q is the input indice. At the end of the training cycle, the weights identify cluster locations, where each weight should be associated with a cluster. An example of this is indicated in Fig.2. Competitive network performance, however, is highly dependent on the number of weights initialized due to the "winner-take-all" nature of the model. Too few weights lead to an oscillation between clusters, while too many weights results in "dead neurons," which are neurons that never win and therefore never update. While these problems do exist in competitive networks, precautions, such as multiple different training sessions with a changing number of weights to confirm results, along with standardization of data, decrease poor performance odds. In this work, competitive networks are employed in junction with the hierarchical clustering. To validate the trends identified in the hierarchical clustering, competitive network models are initialized with a decreasing number of weight vectors, such that strongly-related data still attracts a weight vector. Pre-processing data, such as standardization, ensures that the network is learning the data rather than the numerical values. Moreover, through this, parameters that are not important or misleading in a data set are identified.

E. Supervised Learning

In this work, supervised learning is used in two ways: feature association and time-series forecasting. Similar to feature extraction in the unsupervised learning case, feature association is performed through use of a shallow artificial neural network (ANN). In supervised learning, many neural network architectures exist, from shallow networks with single hyperplanes of $\mathbf{wx} + b$, like a perception model, to deep neural networks and regression neural networks. Shallow neural network models can associate parameters with a target value on a single level, where the weights and bias

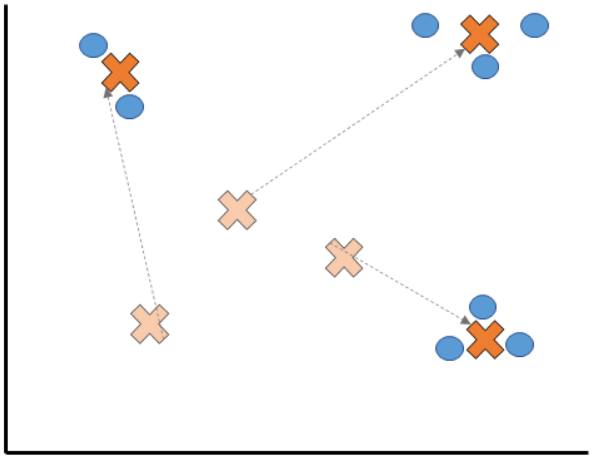


Fig. 2 Competitive Neural Network: initialized and post-training. The weight vectors w, indicated by an X, move to the location of a cluster through an iterative training process, in which a single input is inputted to the network. The weight that is closest to that location will "win," and the weight will update. After the course of successful training, weight vectors associate with clusters in the data set.

hyperplane identify inputs that are most associated with the target. Deeper neural networks are more complex, yielding a high-level of feature association and extraction. For both of these approaches, the model does not remember the last value passed, and the effects of the last value are only identified in the weight update process. On the other hand, regression neural networks (RNNs), retain memory about an input and then use that associated memory value to predict on a time-dependent series [11]. Long-Short-Term Memory (LSTM) models are a variation of the RNN architecture, which was first introduced in [12]. LSTM models, opposed to traditional RNN models, implement preventative measures in error flow causing intra-cell leaky memory, which would prevent a neural network to leverage previous memory values.

III. Approach

An event is defined as the number of CDMs occurring between two specific objects leading to TCA. The earliest predicted miss distances are measured at approximately seven days before TCA. CDMs are considered close approaches if observed within six hours of TCA. Once identified, the CDMs leading up to the close approach are measured via time stamp, spacecraft and mission IDs. One note to make here is that the number of CDMs per event range from two CDMs to over ten CDMs, with most below the ten threshold, as seen in Fig. 3

A. Parameter Identification

For a neural network to learn about an event, enough information must be associated with the item that a network is trying to predict. Therefore, parameter identification is critical. For this research, the following parameters are highlighted for possible use in model training:

1) Miss Distance, m^2: the physical radial distance between Objects 1 and 2, used in collision prediction.
2) Relative Speed, m/s: the speed in which two objects approach relative to each other. Slower speeds are more problematic. [5].
3) LUPI, days: Length of update interval, number of days of observations available to use in orbital determination [5]
4) DC Span, days: Differential Correction, number of days actually used in orbital determination
5) Tracks Used, -: Number of sensor tracks used for orbital determination of the objects [13].
6) RCS, m^2, Radial Cross Section
7) WRMS, -: Weighted Root Mean Square of the residuals from a batch of least squares of the Orbital Determination [13]
8) Bcoef, m^2/kg: Ballistic Coefficient, a measure of a body's ability to overcome air resistance in flight, inversely

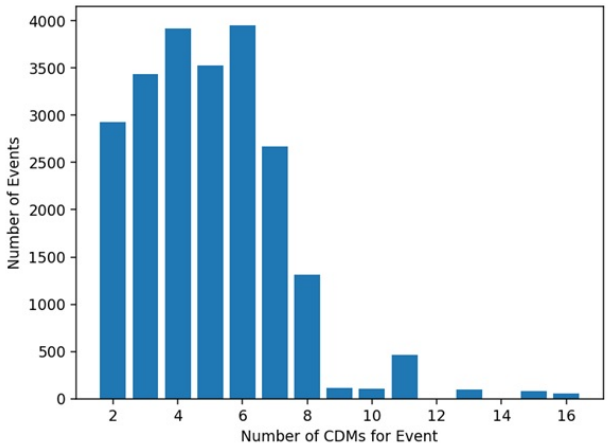


Fig. 3 Overview of CDMs per event for the entire data set. Most events have less than 10 CDMs. When the event is constricted to 5km radial distance, the number of events to train on consist of mostly less than 4 CDMs per event.

related to drag coefficient, cross-sectional area

9) Rcoef, m^2/kg: Solar Radiation Pressure Coefficient, strongly related to AP, known perturbation to satellite
10) EDR, W/kg: Energy Dissipation Rate, closely related to Bcoef, but highly inter-correlated [14]
11) Pc, -: Probability of Collision, encompasses both miss distance and HBR, standard way to define collision risk
12) HBR, m: Hard Body Radius, the sum of both spacecraft circumscribing radii [5]
13) Time to TCA, days: The time until the predicted closest approach between the objects
14) Approach Angle, radians. Using approach angle was based on the success in [15].
15) MHD, m: Mahalanobis Distance, the distance between object 1 and the distribution of the state vector covariances [15, 16]
16) F10, $10^{-22}Wm^{-2}Hz^{-1}$: solar emission flux at 10.7 cm wavelength. [17]
17) AP, -: measure of the general level of geomagnetic activity over the globe for a given (UT) day [18]
18) DST,-: Disturbance Storm Time global index of geomagnetic disturbance

B. Clustering and Competitive Networks

To identify trends with unsupervised learning algorithms, competitive networks and hierarchical clustering are implemented in MATLAB's Statistics and Machine Learning Toolbox [19]. All 300,000+ CDMs are used as inputs to these two models, such that the timeline of an event is not considered in this approach. To limit data loss due to the varying orders of magnitude from the inputs, the data is normalized between the range of $[0, 10]$, with outliers removed to prevent saturation of all values at 0 or 10. For this experiment, $300,000+$ individual CDMs are fed into these unsupervised learning algorithms across various times in observation before TCA. This test is then replicated at events predicted five, three, two and at a half day out until close TCA. First, the hierarchical clustering techniques works from the bottom-up, associating the normalized parameters' neighbors by the closest norm. At the end of the training, a dedrogram is displayed, revealing the closeness of parameter relationship. The competitive networks are initialized with 50, 10, 3, and 2 weights individually and then trained for 10,000 epochs. The results are extracted to identify weight and parameter association.

C. Shallow Binary Classification

Unlike the unsupervised learning techniques methods, an event, as defined earlier, is taken as a single input to both the shallow classification method and the LSTM model. Understanding how Pc values change over the course of the event are critical, especially in understanding how changes in parameter values lead to poor miss distance prediction. To identify which parameter changes are associated with poor predictions and elevated risks of collision, a shallow neural network binary classifier is explored. Two binary classes are selected: intrinsic miss distance standard deviation $md\sigma$, and Pc, such that:

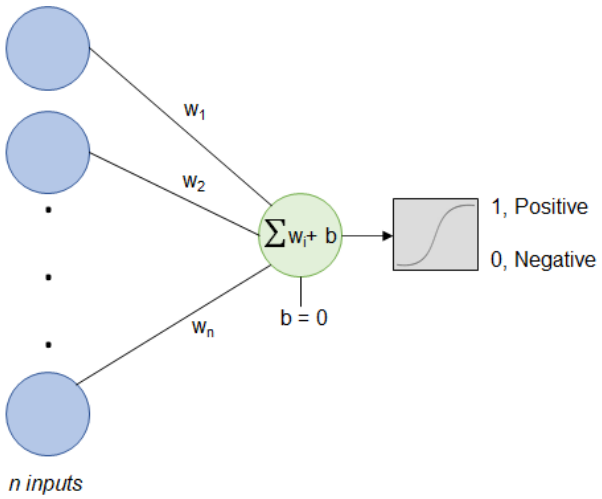


Fig. 4 Shallow Neural Network Binary Classifier with Sigmoid Activation

1) Miss Distance Standard Deviation $md\sigma$, measured by the standard deviation across the event divided by the final miss distance, where $md\sigma >$ average$(md\sigma) \in [1, X]$.
2) Probability of Collision, where $Pc > 10^{-8} \in [X, 1]$.

For this to be an effective interpretation of the parameter importance, the inputs are standardized such that $\mu = 0$ and $\sigma = \pm 1$, and the bias equals zero. Letting the bias remain zero forces the model to completely rely on the weights for classification. This method is explored for all events first and then events that occur at less than 5km miss distance, which reduces the number of available CDMs to approximately 89,000, with 782 events of at least 2 CDMs, and 27 of those events have high risk Pc values. Since Pc values greater that 10^{-8} are far less common than non-worrisome Pc values, careful consideration in what data to use for the binary classifier is important. For instance, if 5% of the events contain Pc values that qualify as threatening to operations, the model may achieve 95% accuracy on inference through learning to predict that no inputs belong to the near-miss classification. This problem is prevented by introducing batch sizes of 1, which prevents the model from generalizing the characteristics of each event, and through reducing the amount of safe approach data, such that a minimum of 25% of training data must contain Pc values greater than the "worry" threshold. For every elevated Pc value, three additional non-threatening events are pulled from the total events data set. For the min length defining an event set to two, the total number of data is 108, drastically reducing the amount of viable data in training. A weighted loss function, which would aid in overcoming the dataset imbalance of high risk events versus safe approaches, is not employed in this work but should be considered for future research. To improve the robustness of the network, an additive white Gaussian noise layer is applied, such that $\mu = 0$ and $\sigma = \pm 0.25$ to the inputs before being passed into the neural network.

D. Long Short Term Memory Neural Network

Since the number of CDMs across an event varies, a mask is applied to each missing time-step. Due to the impertinence of predicting miss distance in short-range events, the LSTM is trained on events that have a miss distance of less than 5km. For the model to learn individual characteristics about an event, the batch size is set to one, and overfitting is prevented through monitoring validation loss across 100-200 epochs with a learning rate of 0.001, using the Adam Optimizer with a mean-squared log error (MSLE) loss function. MSLE is selected over MSE from a trial and error evaluation. The data was split into 80% training, 20% testing, with min-length of an event set to 2 CDMs.

In order to improve accuracy of the LSTM model, four separate LSTM models are constructed with different grouping of parameters. The LSTM leverages the feature selection from the binary classification and the unsupervised methods, utilizing key findings to construct four groupings of the parameters. This method was previously used in [15]. The performance of each of the models is assessed by average error, maximum error, and minimum errors. The miss distance variance across an event is also monitored on path prediction.

The architecture of the LSTM model is illustrated in Table **??**. The masking layer indicates the time steps that the model should ignore in training. Two LSTM layers, activated with ReLU, were selected based on an accuracy

Table 1 LSTM Model Architecture with output dimensions illustrated.

Layer	Output Dimensions
Masking	(MaxLen, Features)
LSTM +ReLU	(5*MaxLen)
Repeat Vector	(1, 5*MaxLen)
LSTM + ReLU	(5*MaxLen)
Time Dis. Dense + Linear	(Pred. Miss Distance)

performance analysis with [0-5] LSTM hidden layers. Overfitting occurs rapidly beginning at three LSTM layers. Finally, a time distributed dense layer, applies a $\mathbf{wx} + b$ to each of the time steps, outputting the predicted miss distance.

IV. Results

A. Parameter Identification

In Fig. 5, the clustering results of four different competitive neural networks, initialized at 50, 10, 3, and 2 classes respectively, are illustrated. At 50 available classes, several parameters— Rcoef (pri/sec), EDR (pri), MHD, RCS (sec), AP— are still grouped together with miss distance and Pc, which suggests that there is a strong correlation between these parameters. As the number of available classes decreases, other parameters— Bcoef (pri), Tracks Used (sec), WRMS (sec), RCS (pri), Rcoef (sec), F10— join the grouping with Pc in Group 1. Group 2 includes the approach angle, relative speed, HBR, DST, Time, DC Span, and LUPI numbers, which are expected to be grouped together due to their relationship to time. While Fig. 5 focuses the entire data set, the same experiment at events predicted at five, three, two, and half day out steps is performed.

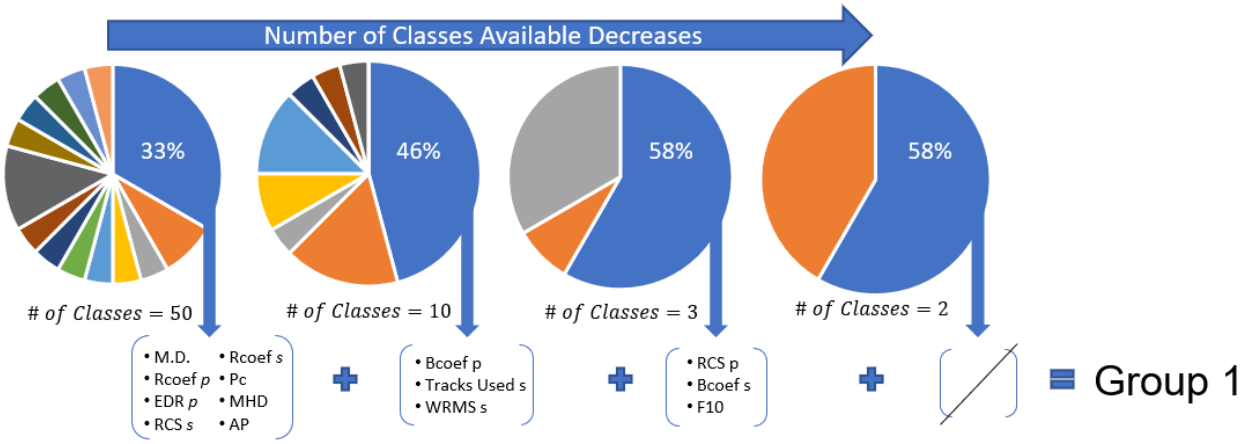


Fig. 5 Competitive Network grouping for close approach events. Strong correlations are observed with lagniappe weight initialization, such that 33% are still grouped together.

In tandem with these relationships, the hierarchical clustering results share similar results. Miss Distance and MHD are closely related, and as the level of closeness decreases, a similar groping emerges as identified in Fig. 5. As expected, MHD, Pc, and miss distance are most closely related together, and this is due to the nature of Pc's relationship mathematical relationship with miss distance and the covariance that MHD measures. Since Pc encompasses MHD and miss distance in its calculation, these relationships are expected, along with being indicative that the hierarchical clustering and competitive networks were achieving the predicted results. The inclusion of Rcoef, EDR, RCS, and AP metrics indicated that these values are related. Solar radiation pressure is a known perturbation, and the effects of solar radiation pressure have been observed to corrupt orbital projections [20]. In [21], large energy dissipation rates are

associated with increased propagation errors as well. AP values, as further assessed in [22], indicates that geomagnetic indices are known to cause errors in orbital determination as well.

B. Binary Classifier

The shallow binary classifier for the $md\sigma$ model achieves an overall accuracy of 85.8%, with a sensitivity of 85% and specificity of 86.5%. For $md\sigma$, the network identifies seven parameters that affect miss distance predictions: LUPI (pri), Tracks Used (pri), EDR (pri), Tracks Used (sec), EDR (sec), Time to TCA, and MHD. The strongest relationships emerge with MHD, EDR (sec), Time, and Tracks Used (sec). The network determines that below average MHD values correlate with $md\sigma$, which suggests that covariance realism is affecting the ability to predict miss distance effectively.

For high risk classification, the best performing model achieves 94.4% accuracy with a 100% sensitivity, with lower tier models performing at about 70% sensitivity. Sensitivity is an important metric for analyzing risk prediction, for higher sensitivity values indicated a low number of false negatives. In risk mitigation, a false positive in which an unnecessary maneuver is taken is safer than a false negative, in which a maneuver is not taken and the risk results in disaster.

To compare the $md\sigma$ and Pc weight associations, EDR (sec), Time, and Tracks Used (sec) are strongly related to the identification of high risk events in the same way that these parameters are related to high $md\sigma$ events. Lower than average tracks used and EDR for the secondary object, along with above average time (several days out) are related to high risk events and high $md\sigma$ events. Other correlations identified with high risk events are large EDR (pri) and above average F10.

C. Miss Distance Prediction

Table 2 The four groups constructed to test the effectiveness of the LSTM model, along with determining which of the groups minimized miss distance error for events with less than 5k approaches. Group 4 was not tested with close approaches due to a poor performance on the overall approach data set.

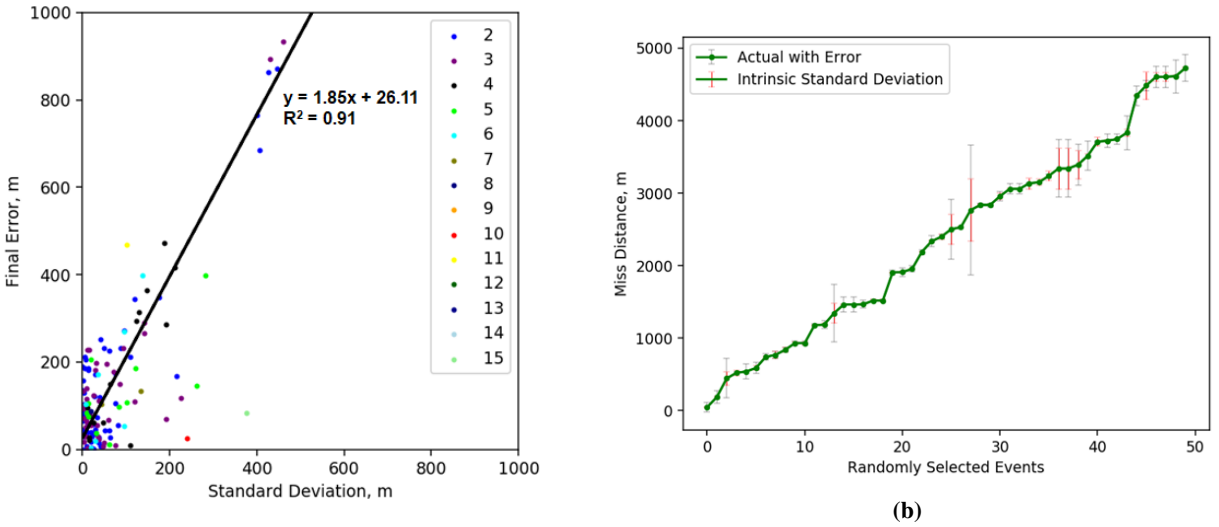
Group	Miss Distance	Relative Speed	LUPI (pri)	Tracks Used (pri)	RCS (pri)	WRMS (pri)	BCoef (pri)	Rcoef (pri)	EDR (pri)	LUPI (sec)	Tracks Used (sec)	RCS (sec)	Bcoef (sec)	Rcoef (sec)	EDR (sec)	Pc	HBR	Time	Angle	MHD	F10	AP	DST
1	x						x	x	x		x	x	x	x	x	x		x		x	x	x	
2	x	x	x	x	x		x	x	x	x	x	x	x	x	x		x				x	x	x
3	x	x	x	x	x				x	x		x			x		x				x	x	x
4	x		x			x				x	x			x		x	x		x		x		

Group 2, indicated in Table 2, outperformed the other groupings for miss distance prediction. The following results were tested using the parameters selected in Group 2. Figure 6 displays the LSTM's performance for fifty randomly-chosen predicted events, where the green line represents the true miss distance and the grey represents the predicted error range by the LSTM model. The model does not appear to perform better at a particular miss distance range. However, as the intrinsic standard deviation $md\sigma$ increases, indicated by the red bars, the error on prediction doubles. This relationship is further illustrated in Fig. 6. Moreover, events that have a smaller number of CDMs per event typically run above the variance line, and while events with a larger amount of information to train from are below. The average time step before prediction was 1.8 days to TCA.

Table 3 Resulting errors from LSTM model performance, with Group 2 selected for further evaluation.

Parameter Groups	1	2	3
Average Error, m	407	211	230
Max Error, m	3900	2480	2400
Min Error, m	1.2	1.6	0.98

A major compromise in training the LSTM model is the trade-off between having a large number of events to train on and having a minimum number of CDMs in a given event. Due to a lack of substantial (thousands of) data for events that contain more than three CDMs in the 5km range with a close TCA approach, fine-tuning the model is challenging, and overfitting is a common problem with increased network depth. However, the model averages about 0.2km of error per testing set. Removing events with intrinsic high $md\sigma$, the average error drops to below 0.1km. Increasing the number of CDMs per event and increasing the number of high risk events in a dataset would encourage the model to navigate towards lower prediction error by extending the amount of training examples before overfitting.



(a) **(b)**

Fig. 6 This model was trained on Group 2 parameters, using events with at least two CDMs. In (a), the relationship between $md\sigma$ and prediction error is further illustrated. The strong linear relationship, denoted by the R^2 value, indicates that for each $md\sigma$ meter, the final error nearly doubles. With very low $md\sigma$, the model can predict the miss distance with base error of 26.11m. The number of CDMs in a given event is depicted to the right. Fig. 8 (b) illustrates fifty randomly selected events' prediction steps from the LSTM model, with the average prediction step taken at 1.8 days before TCA. Grey error bars represent the prediction error, while the red error bars denotes the intrinsic miss distance variance over the time steps $md\sigma$.

V. Conclusions and Future Work

Using a combination of unsupervised and supervised learning techniques, additional parameters, such as EDR and Rcoef, are associated with increased miss distance prediction and collision risk assessment. Moreover, $md\sigma$, a measure of what the intrinsic standard deviation of the miss distance across an event is, illustrates a model's limitation in the final predication step. If an event has high $md\sigma$, miss distance prediction is subject to error; however, using parameter-set analysis in determining which events are subject to high $md\sigma$ helps identify confident prediction steps. Accounting for perturbations may aid in the prediction step, and an understanding the reliability of covariance measure is also necessary to identify the weak points in predication.

Future iterations of this work includes developing a weighted loss function, which would increase the number of data used for training by weighting the high risk events to have an impact on model training without partitioning. A concatenation model, such that the LSTM model learns to identify $md\sigma$, would account for that in a confidence value in each predicted miss distance. Finally, increasing the number of available data would improve the overall predictive qualities of these models.

VI. Acknowledgments

Robertson would like to thank her mentor, Dr. Alinda Mashiku, for the opportunity to work at Goddard Space Flight Center in Summer 2020, despite the COVID-19 Pandemic environment. Moreover, Robertson would also like to thank the NASA CARA Team for the opportunity to learn and contribute.

References

[1] Newman, L. K., Frigm, R., and McKinley, D., "It's Not a Big Sky After All: Justification for a Close Approach Prediction and Risk Assessment Process," *AAS/AIAA Astrodynamics Specialist Conference*, 2009. Available: `https://ntrs.nasa.gov/citations/20090032050`

[2] "As SpaceX Launches Dozens Of Satellites At A Time, Some Fear An Orbital Traffic Jam," *NPR*, 2019. Available: https://www.npr.org/2019/11/11/778219787/as-spacex-launches-dozens-of-satellites-at-a-time-some-fear-an-orbital-traffic-j

[3] Peterson, G., "Space Traffic Management in the Age of New Space," *Center for Space Policy and Strategy*, 2018. Available: https://aerospace.org/sites/default/files/2018-05/SpaceTrafficMgmt_0.pdf

[4] "Space Debris and Human Spacecraft," *NASA.gov*, 2017. Available: https://www.nasa.gov/mission_pages/station/news/orbital_debris.html

[5] Hejduk, M. D., and Johnson, L. C., "Evaluating Probability of Collision (Pc) Uncertainty," NASA GSFC-E-DAA-TN30990, April 2016. Available: https://ntrs.nasa.gov/citations/20160005313

[6] Ghrist, R. W., Plakalovic, D., "Impact of Non-Gaussian Error Volumes on Conjunction Assessment Risk Analysis," *AIAA/AAS Astrodynamics Specialist Conference*, 2012. Available: https://doi.org/10.2514/6.2012-4965

[7] Yang, J., "Visual Hierarchical Dimension Reduction for Exploration of High Dimensional Datasets," *Symposium on Visualization, VisSym*, 2003.

[8] Gilpin, S., Qian, B., Davidson, I., "Efficient hierarchical clustering of large high dimensional datasets," *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pp. 1371-1380, 2013. DOI: 10.1145/2505515.2505527

[9] Steinbach, M., Karypis, G., Kumar, V., "A Comparison of Document Clustering Techniques," *University of Minnesota, Technical Report #00-034*, 2000.

[10] Hagan, M. T., Demuth, H. B., Beale, M. H., and Jesús, O. D., "Neural Network Design," 2nd ed., 2014. Available: https://hagan.okstate.edu/NNDesign.pdf

[11] Specht, D. F., "A General Regression Neural Network," *IEEE Transactions on Neural Networks*, Vol. 2, No. 6, pp. 568-576, 1991. DOI: 10.1109/72.97934

[12] Hochreiter, S. and Schmidhuber, J., "Long Short-Term Memory," *Neural Computation*, Vol. 9, No. 8, pp. 1735-178-, 1997. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[13] Conjunction Data Message, The Consultative Committee for Space Data System Standards, 2013.

[14] Storz, M. F., Bowman, B., and Branson, J., "High Accuracy Satellite Drag Model (HASDM)," presented as paper AIAA-2002-4886 at the AIAA/AAS Astrodynamics Specialist Conference and Exhibit, Monterey, California, August 5–8, 2002

[15] Mashiku, A., Frueh, C., Memarsadeghi, N., Gizzi, E., Zielinski, M., Burton, A., "Predicting Satellite Close Approaches Using Statistical Parameters in the Context of Artificial Intelligence," *Astrodynamics Specialist Conference*, 2019. Available: https://ntrs.nasa.gov/citations/20190029019

[16] Joseph, E., Galeano, P., Lillo R. E., "The Mahalanobis Distance for Functional Data With Applications to Classification," *Technometrics*, Vol. 57, Iss. 2, pp. 281-291, 2015. Available: https://doi.org/10.1080/00401706.2014.902774

[17] "10.7cm Solar Radio Flux," *NorthWest Research Assiciates, Inc*, 2020. Available: https://spawx.nwra.com/spawx/f10.html#:~:text=The%20F10.,DRAO)%20in%20Penticton%2C%20Canada.

[18] Faser-Smith, A. C., "The Spectrum of the Geomagnetic Index Ap," *Stanford Electronics Laboratories*, 2017.

[19] The Mathworks, Inc. Cluster with a Competitive Neural Network. R2019b. Available: https://www.mathworks.com/help/deeplearning/ug/cluster-with-a-competitive-neural-network.html

[20] Melbourne, W.G., Navigation between the Planets, *Scientific American*, Vol. 236, No. 6, pp. 58-76, 1975. Available: https://www.jstor.org/stable/24950372

[21] Hejduk, M. D., and Ghrist, R. W., "Solar Radiation Pressure Binning for the Geosyncronous Orbit," *AIAA/AAS Atrodynamics Specialist Conference*, 2011. Available: https://ntrs.nasa.gov/citations/20110015238

[22] Warner, J. G. and Lum, A., "On the Mitigation of Solar Index Variability for High Precision Orbit Determination in Low Earth Orbit," *US Naval Research Laboratory* 2016. Available: https://apps.dtic.mil/sti/pdfs/AD1013717.pdf