Detecting Network Anomalies with Transfer Learning and Support Vector Machines

Vincent Terrelonge
Advisor: Dr. Abdelfattah Amamra
Department of Computer Science, California State Polytechnic, Pomona, Cybersecurity Awareness Fair 2024

College of Science
Computer Science Department

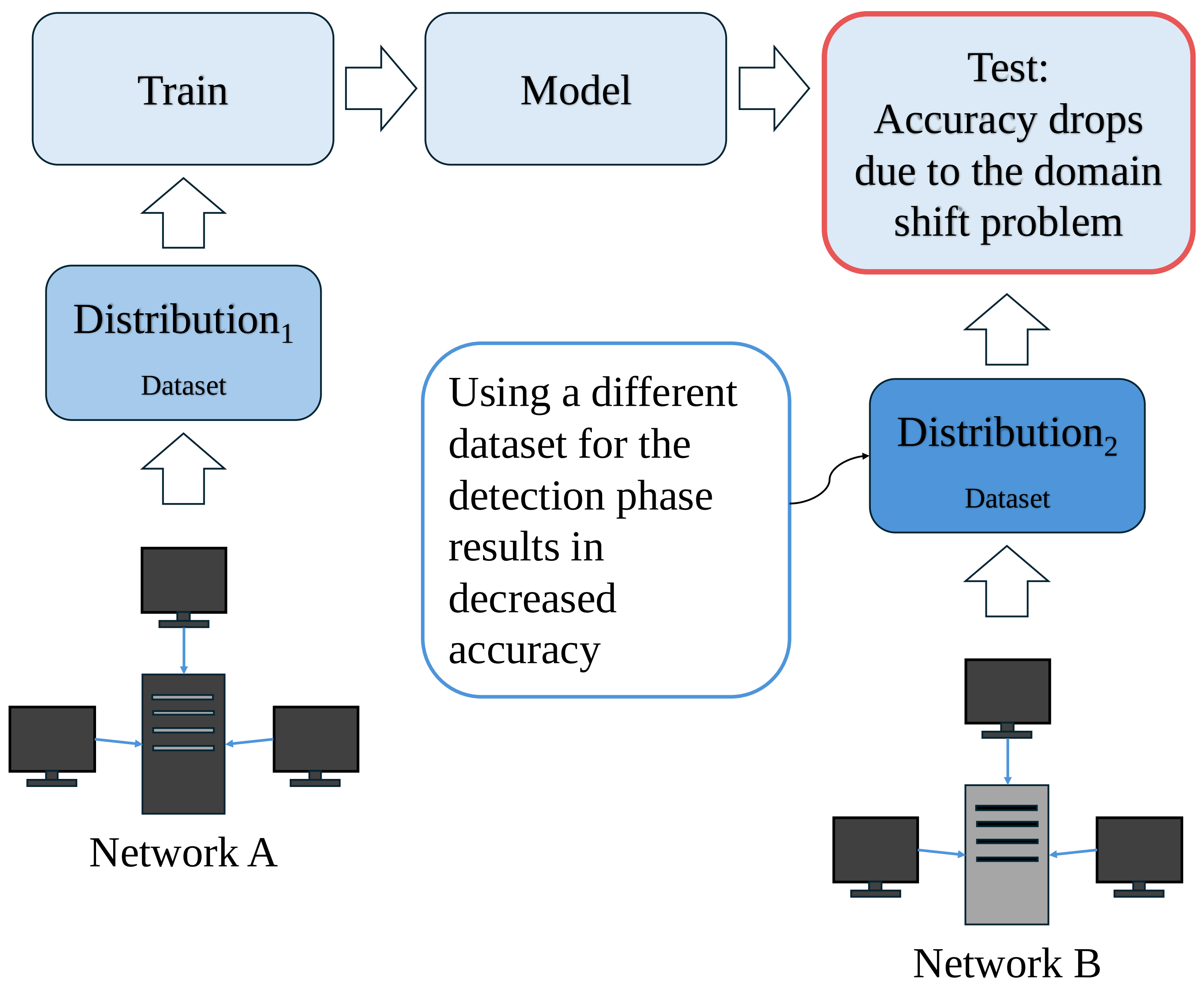
Problem

The problem with Supervised Machine Learning

Supervised machine learning (ML) models are widely used to tackle the problem with intrusion detection; however, these approaches face limitations:

- Large amounts of labeled data are required for supervised ML models.
- Real world datasets are limited due to privacy and security policies.
- Synthetic datasets are prevalent and are unable to generalize well.
- Supervised ML models suffer from the domain shift problem [1].
- Detecting new attacks are unlikely without new features or retraining.

Figure I. Supervised Machine Learning Framework



Train

Model

Test: Accuracy drops due to the domain shift problem

Distribution$_1$ Dataset

Using a different dataset for the detection phase results in decreased accuracy

Distribution$_2$ Dataset

Network A

Network B

Proposed Solution

Transfer Learning

Transfer Learning (TL) will allow us to overcome the limitations of supervised (ML) by transferring the knowledge of the source domain to the target domain. TL helps us address the challenges such as data scarcity and the domain shift problem [1].

Benefits of Transfer Learning

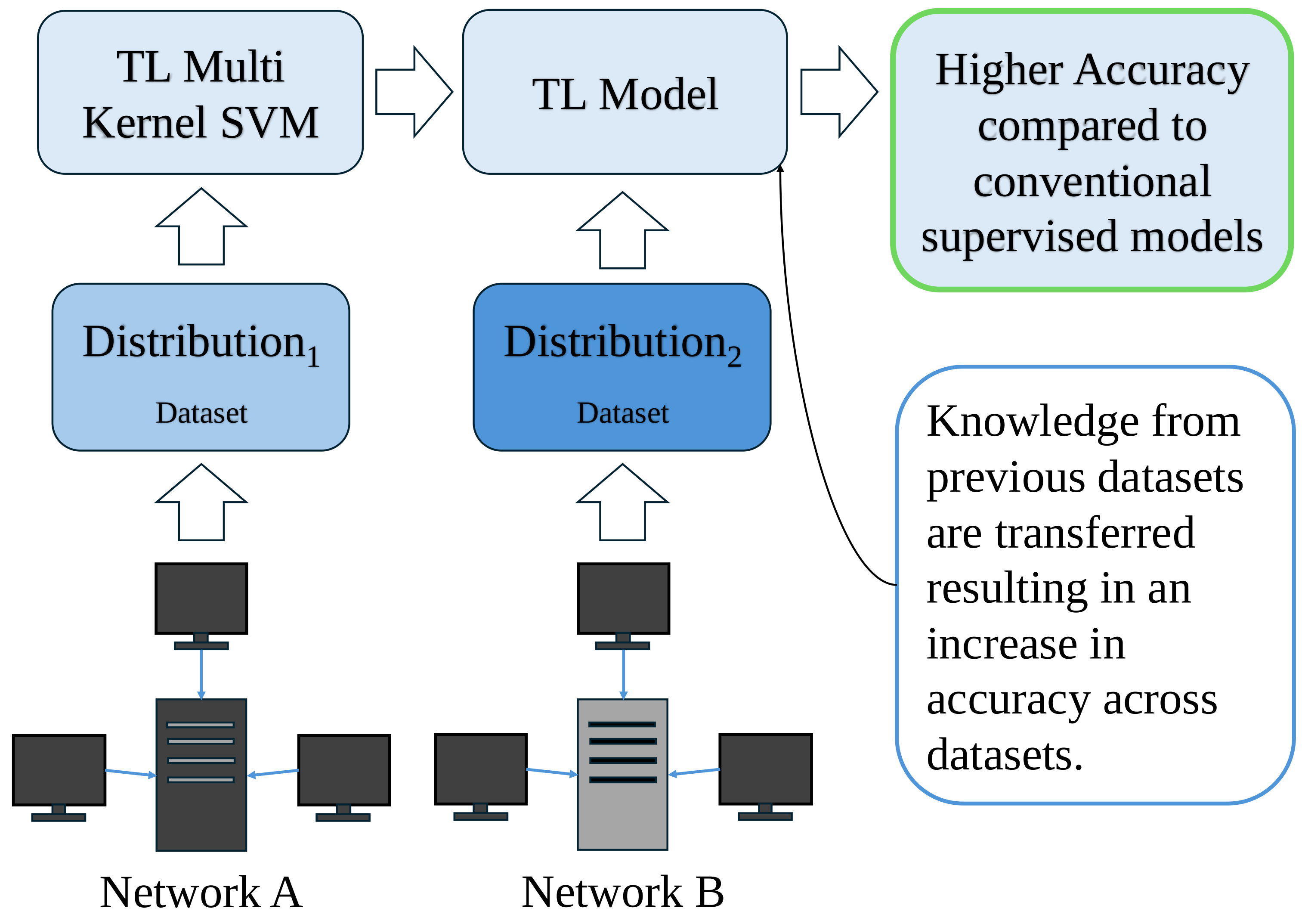
- Transfers knowledge across domains.
- Reduces the amount of labeled data required.
- Reduces training time.
- Enhances performance in target domains.

Proposed Solution Models

Transfer Learning Framework

We will use a large dataset labeled Distribution$_1$ to train our model. For a new dataset label Distribution$_2$, we will transfer the knowledge or weights to the new iteration. New iterations correspond to new datasets with transferred weights.

Figure II. Transfer Learning Framework



TL Multi Kernel SVM

TL Model

Higher Accuracy compared to conventional supervised models

Distribution$_1$ Dataset

Distribution$_2$ Dataset

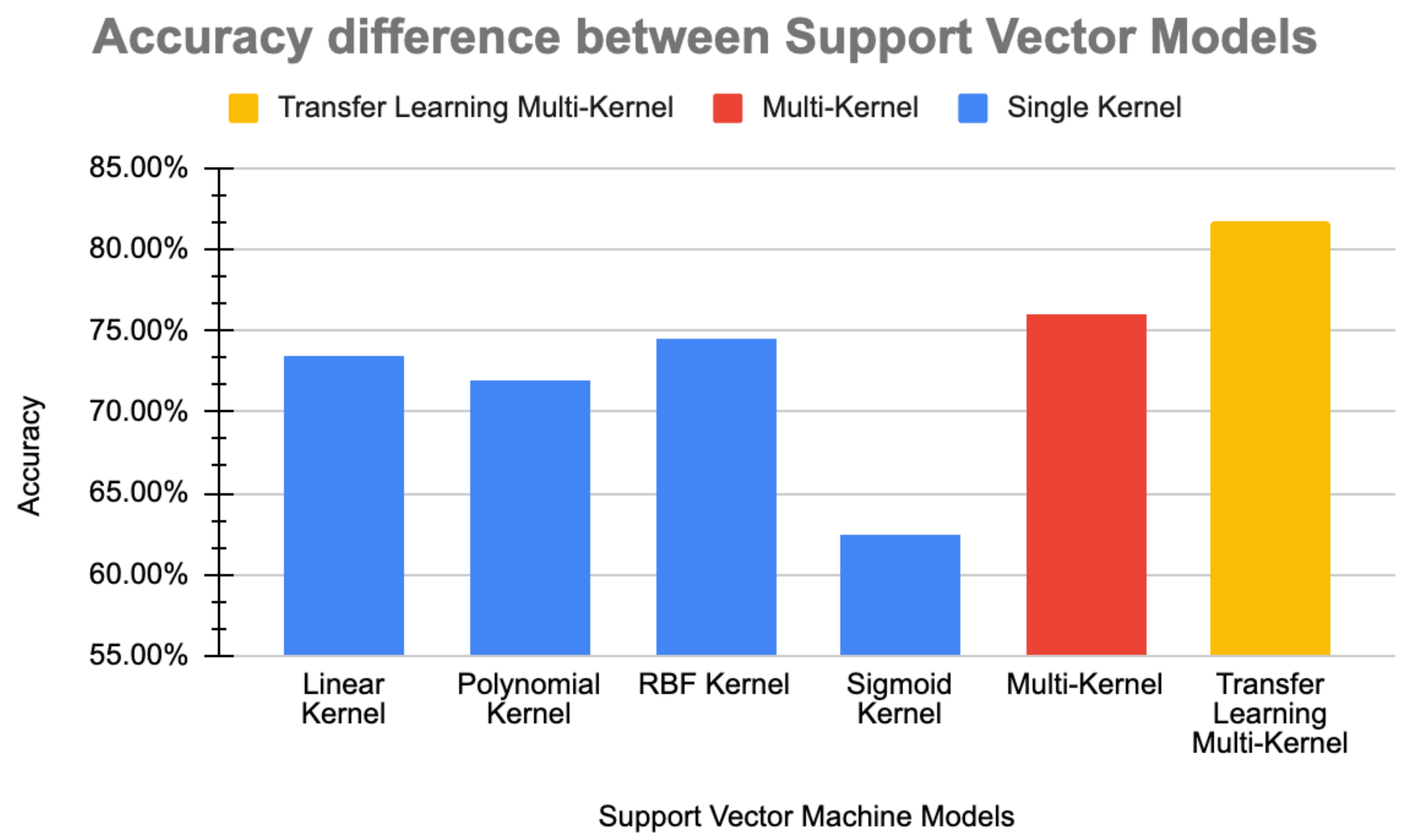
Knowledge from previous datasets are transferred resulting in an increase in accuracy across datasets.

Network A

Network B

Experimental Results

Figure III. Proposed Algorithm Results

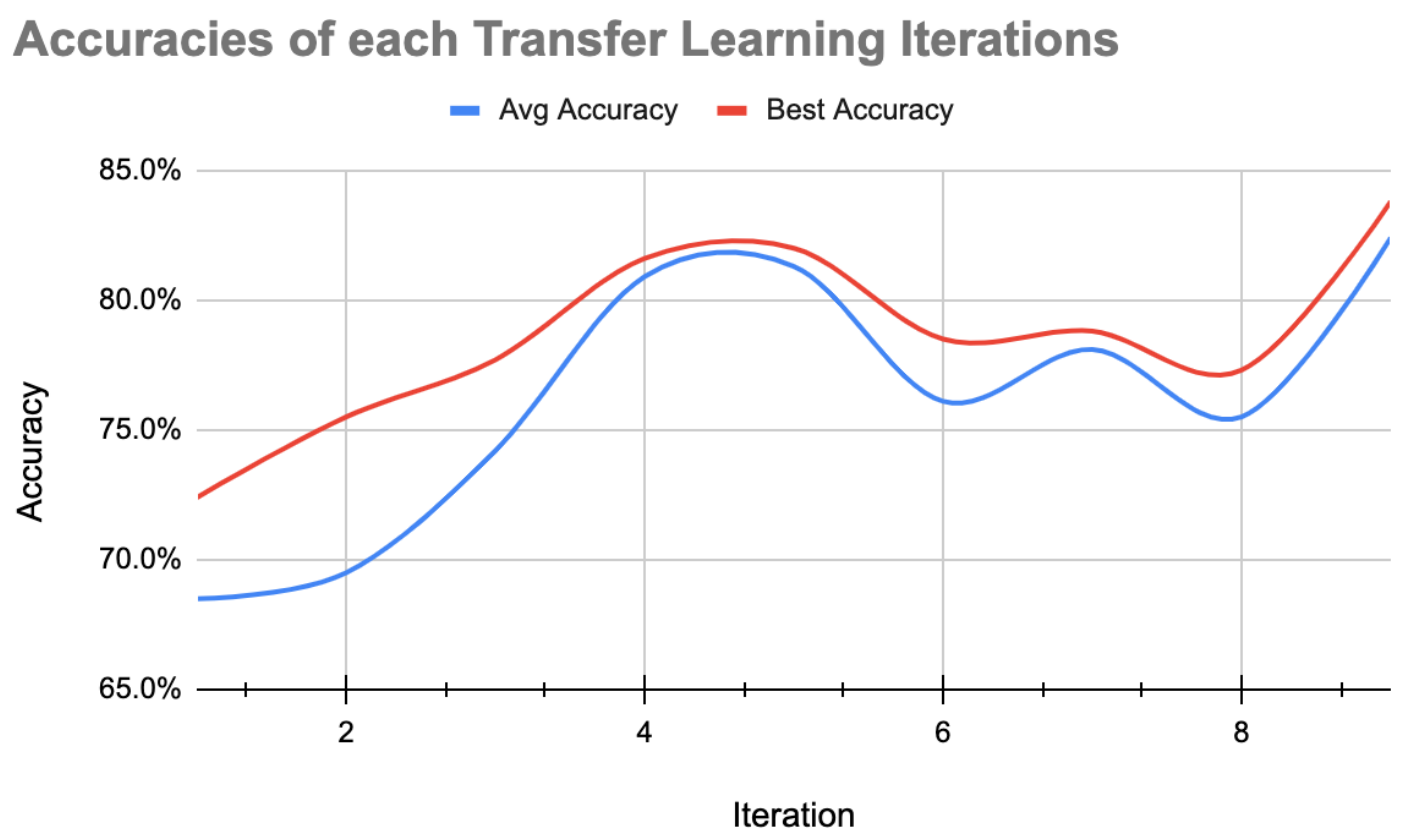


Experimental Results

Transfer Learning Multi-Kernel SVM Results

Figure IV. shows an increase in average and best accuracy with the transfer learning multi-kernel SVM model. The overall accuracy increased with 9 transfer iterations.

Figure IV. Avg and Best TL Accuracies Over 9 Iterations



In each iteration, the average accuracy across 100 weight optimizations is nearly the same as the best accuracy for that transfer iteration. This suggests the model generalizes well since the average accuracy consistently approaches the optimal performance.

Conclusion and Future Work

Transfer learning and multi-kernel support vector machines give a promising new foundation in intrusion detection. Transfer learning MKSVM algorithms are expected to outperform conventional supervised ML techniques in detecting network anomalies. For future experiments, it would be worth testing how the percentage of target data in the source dataset affects the overall accuracy. It would also be worth testing how a deterministic MKSVM affects the accuracy.

References

[1] S. Ma et al., "Deep Into the Domain Shift: Transfer Learning Through Dependence Regularization," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2023.3279099.

[2] Zafar Iqbal Khan, Mohammad Mazhar Afzal, and Khurram Naim Shamsi. “A Comprehensive Study on CIC-IDS2017 Dataset for Intrusion Detection Systems”. In: 2024 *International Research Journal on Advanced Engineering Hub (IRJAEH).*

[3] Noah Reef. “Stochastic Domain Transfer Multiple-Kernel Boosting with Application to Anomaly Detection in Encrypted Network Traffic”. In: 2022 *California State Polytechnic University, Pomona*