Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Decoding Crude Oil Trade Link Dynamics: Evidence from Machine Learning and Deep Learning Models

Ritu Singh

ritusingh20@iitk.ac.in

Indian Institute of Technology Kanpur

Research Article

Keywords: Crude oil, Link prediction, Machine learning, Deep Learning models, Dynamic trade networks, Energy trade

Posted Date: April 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8787165/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Decoding Crude Oil Trade Link Dynamics: Evidence from Machine Learning and Deep Learning Models

Ritu Singh[a*ψ]

[a] Department of Economic Sciences, Indian Institute of Technology Kanpur, Uttar Pradesh, 208016, India

Abstract

Globally, the crude oil energy landscape is influenced by supply and demand dynamics, energy security considerations, and volatile international relations. This research proposes a computational framework for predicting the formation and dissolution of bilateral crude oil trade links with specific reference to major crude oil exporting countries by employing six classification-based machine learning and deep learning algorithms as empirical tools. It harnesses a dataset covering two decades, comprising node-specific and edge-specific features to capture the economic, historical, and geopolitical nuances affecting crude oil trade linkages. Results indicate that non-linear interactions between geopolitical alignment, economic size, and historical trade inertia play a decisive role in shaping crude oil trade link dynamics. Out of all models employed, the Random Forest Classifier can aptly predict both active and inactive crude oil trade links, outperforming linear and deep-learning models and further emphasizing the immense potential of employing ML algorithms in economics. Moreover, the study has been further enhanced by incorporating network-based measures such as link turnover rate, substitution index, reactivation ratio, and persistence ratio that are pivotal in uncovering systemic shifts and contagion effects in global crude oil markets. These findings yield critical insights for policymakers by enabling the anticipation of shifts in trade linkages, thereby facilitating the development of adaptive strategies to address evolving global challenges. In doing so, the study also contributes to the advancement of prediction-oriented methodological frameworks in the domain of energy security.

Keywords

Crude oil, Link prediction, Machine learning, Deep Learning models, Dynamic trade networks, Energy trade

Classification codes

F170 [Trade: Predicting and Simulation], F14 [Country and Industry Studies of Trade], F51 [International Conflicts], Q4 [Energy]

1. Introduction

This paper aims to predict the formation and dissolution of crude oil trade links of major oil-exporting countries. In this regard, this study has incorporated a myriad of variables, as node-based and edge-based features, which can capture historical, economic, and geopolitical perspectives of dynamic crude oil trade relationships. There is evidence that crude oil is

* Corresponding author.
E-mail address: ritusingh20@iitk.ac.in, ORCID: https://orcid.org/0009-0006-8274-7870

ψ A smaller version of this paper has been published in the proceedings of the 2024 IEEE International Conference on Future Machine Learning and Data Science (FMLDS). https://doi.org/10.1109/FMLDS63805.2024.00025

intricately entangled with geopolitics and, hence, is very much affected by evolving or degrading global relations.

The global energy landscape is dominated by crude oil. Given the asymmetric distribution of this crucial energy reserve, along with the nuances of demand and supply dictated by the global oil market, countries are eager to ensure a steady and secure supply of crude oil. A nation's energy security landscape is heavily dependent on safe and easy access to crude oil, and in turn, plays a decisive role in the strategic decisions of countries on the world stage.

Crude oil consumption increased to 97.3 million barrels/day in the year 2022, according to the Statistical Review of World Energy (2023) report. The document further emphasized that in 2022, North America, South & Central America, Europe, the Commonwealth of Independent States (CIS), the Middle East, Africa, and Asia Pacific respectively asserted 26.9%, 6.8%, 3.3%, 14.9%, 32.8%, 7.5%, and 7.7% share in the worldwide production of crude oil with the corresponding shares in consumption of crude oil being 24.2%, 6.3%, 14.5%, 4.8%, 9.7%, 4.3%, and 36.3% respectively. The imbalance between the demand and supply of crude oil is explicitly demonstrated by these statistics.

An intricate interplay of strategic, geopolitical, and economic factors heavily influences the formation and dissolution of crude oil trade links. Cappelli et al. (2023) ascertained that far-reaching difficulties are being endured by oil-deficient countries in a bid to attain an equilibrium between the energy facets concerning energy accessibility, energy security, and environmental sustainability. The unique thing about oil is that it has unequal distribution and scarce availability, inducing nations to exercise oil control as an instrument to achieve global supremacy (Wakeford & Swilling, 2014; Colgan, 2013; Su et al., 2019). This oil power dominance and utilization have resulted in a nexus of international military confrontations and conflicts, for instance, the Iran-Iraq war, the Gulf War, and the Libyan War. All of these events had substantial repercussions on oil prices.

For a nation to be energy independent, the key lies in safeguarding the supply chain, given the fact that dependence is often disproportional and unbalanced. Moreover, dependence on vulnerable trade partners can result in supply turmoil and unstable prices. Further, exploiting energy resources as a political apparatus can endanger the well-being of a country. A noteworthy instance is the imposition of an embargo by the Organization of the Petroleum Exporting Countries (OPEC) on oil exports to Japan, Canada, the Netherlands, the U.K., and the USA during the 1970s. This was accompanied by a reduction in production and an increase in prices. The outcome was economic hardships, stock market crashes, and oil shocks. The after-effects provoked abundant efforts by oil-deficient nations to make crude oil apolitical, along with alterations in subsequent strategies to ensure energy self-sufficiency (Skeet, 2014; Shaffer, 2011; Kanchana et al., 2016).

Recent developments have led to the emergence of Asian countries such as India and China as chief importers of crude oil, thus establishing them as crucial players in the crude oil trade linkages. As such, the endeavour to achieve energy security is the main motive behind establishing alliances by oil-importing nations, which are more often than not based on complex conditions. This reasoning has been well explained by Colgan (2013). He contends the simplistic notion of direct resource conflicts involving countries waging war just to gain control over oil reserves. Instead, he explains a full-spectrum framework that includes a myriad of indirect ways in which crude oil contributes to international disputes and

confrontations. As per the author, oil acts as the main catalyst for armed conflicts in the present time. This is not only because of the resource-based warfare, but mainly owing to the enigmatic convergence of strategic, political, and economic motivations provoked by oil.

Against this backdrop, there is an augmenting need to develop empirical frameworks that are able to predict changes in dynamic and volatile trade link structures. The first contribution of this paper is to the literature on international energy trade by examining the conditions under which crude oil trade relationships form, or dissolve. In this regard, this study brings a prediction-oriented methodological advancement via a novel application of machine learning and deep learning modelling techniques and the subsequent temporal validation. This framework predicts not only the formation but also the dissolution of crude oil trade links, which is a critical yet underexplored dimension of energy trade modelling. This configuration directly enables the early identification of fragile relationships, thus offering subtle insights to both oil-exporting and oil-importing countries. The task has been formulated as a spatio-temporal classification problem.

The second contribution relates to the emphasis on node-specific and edge-specific attributes for prediction. This methodology proves especially beneficial in scenarios involving dynamic networks characterized by evolving network structures. A noteworthy aspect is that it does not necessitate pre-existing familiarity with network-centric characteristics and hence is quite general. The insights from economic, political, semantic, and historical features have been integrated with network-derived structural measures such as link turnover rate, substitution index, reactivation ratio, and persistence ratio, which bridge the gap between gravity-style models and the burgeoning literature on trade network evolution.

The third and final contribution is empirical through a two-decade dyadic unbalanced panel dataset on global crude oil trade linkages for major crude oil exporting countries. A unique, unbalanced panel dataset (2002-2021) has been constructed and analyzed in this research that captures the long-run longitudinal trade link dynamics of oil-exporting countries. The dataset allows this study to opt for granular exploration of link formations, dissolutions, and reactivations, unlike previous studies that focus solely on active links or rely on static bilateral values.

Overall, the proposed framework in this research underscores the merit of computational prediction tools in comprehending nonlinear, relational, and temporal patterns in crude oil trade dynamics. Moreover, it demonstrates the applicability of Machine Learning (ML) and Deep Learning (DL) models beyond conventional predicting contexts like price prediction or macroeconomic indicators. Therefore, this paper, in terms of its contribution, essentially lies at the juncture of predictive modelling, computational Economics, and trade network analysis. The following sections comprehensively elaborate on the existing literature on the use of ML and DL models in Economics and energy-related prediction, the dataset and variables employed, methodology, experimental results, comparative analysis of models' performances, and the in-depth analysis of network-based metrics. The paper concludes with implications and future research directions.

2. Literature Review

There has been an increasing number of research papers in recent times incorporating machine learning techniques in Economics. This has been reviewed in detail by Ghoddusi,

Creamer & Rafizadeh (2019). The authors emphasize that ML techniques are in use in areas of crude oil price prediction, electricity price prediction, predicting prices of other energy commodities, predicting/modelling energy consumption and demand, model calibrations (identifying parameters in energy demand-supply systems), developing trading strategies, estimating the structure and dynamics of energy systems, and policy analysis. The algorithms include Artificial Neural Networks (ANNs), SVM, hybrid models, ensemble methods, feedforward and feedback multi-layer networks, and fuzzy logic, to name a few.

Creamer & Rafizadeh (2019) further explain that employing ML models has both pros and cons. The models provide superior prediction accuracy and, in many scenarios, outperform traditional econometric models, can handle large and heterogeneous data, uncover complex relationships, and provide flexibility in applications. The limitations often include the risk of overfitting, black box nature (lack of transparency), large data requirement, and lack of statistical inference. To counter these challenges, this research has employed regularization techniques, focused on careful feature selection, maintained model complexity at par with the size of the dataset, and employed cross-validation measures.

For intricate predictive tasks in Economics in which statistical and econometric models struggle, DL has emerged as a transformative approach. Relative to traditional ML models requiring manual feature engineering and constrained by tree-based logic or linearity, more complex DL models offer greater representational power and hierarchical feature abstraction (Mullainathan & Spiess, 2017). Advanced computational models have been categorized into four types by Nosratabadi et al. (2020) that involve deep learning, hybrid deep learning, hybrid ML, and ensemble methods with separate and unique upsides depending on analytical goals and the data structure under consideration.

The majority of the DL implementation in Economics is for price-predicting tasks. A transfer learning-enhanced LSTM framework has been used by Cen and Wang (2019) to comprehend short-term fluctuations in crude oil prices. Combining sentiment analysis and topic modelling with a Convolutional Neural Network (CNN), Li et al. (2019) have tried to predict price movements. Similarly, Particle Swarm Optimization (PSO) has been used to optimize LSTM along with their integration with sentiment variables for predicting crude oil market volatility by Jiao et al. (2021). Alternatively, hybrid DL models combining decomposition methods such as Ensemble Empirical Mode Decomposition (EEMD) and Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) with LSTM and Gated Recurrent Unit (GRU) have been employed by Wu et al. (2023) and Deng et al. (2021) for better accuracy and robustness.

Another study by Shen et al. (2024) employs a CNN-BiLSTM-Quantile Regression pipeline in order to probabilistically predict crude oil prices, demonstrating the capability of DL algorithms in dynamic environments. Intricate models like GRU-CNN (Han et al., 2015) and transformer-based temporal predicting (Liang et al., 2023) have shown promising results in dynamic environments, but these algorithms are often data-intensive, thus making them unsuitable for medium-scale panel datasets. Similarly, Zheng et al. (2023) argue for context-appropriate model selection in order to achieve a careful balance between complexity and interpretability as far as economic prediction tasks are concerned.

Upon considering all the DL architectures, CNNs and ANNs stand out owing to their adaptability across various domains. CNNs are most apt when it comes to the extraction of

spatial or pattern-based relationships, given the fact that this model is increasingly being employed to structured panel and sequential datasets in spite of their initial implementation in image-based data. Turning to the ANN model, it is relatively simpler in architecture but provides a robust baseline for classification tasks and has often been used in the literature as a foundational model prior to transitioning to more complex algorithms.

Two major considerations motivate the selection of these two models in this research. The first reason relates to the nature of the dataset, which is structured, dyadic, and moderately sized, unbalanced panel data spanning over two decades, and given this nature of the dataset, it doesn't readily support highly parameterized models like Long and Short Term Memory (LSTM) or Transformer networks requiring larger training data and longer sequences to perform well. Secondly, the aim of this paper is to analyze whether spatial channel encoding (in CNN) and nonlinear abstraction (in ANN) can lead to an improvement in the predictive power relative to traditional ML models, which offer a clean and computationally efficient starting point. The logic follows from the fact suggested by Goodfellow et al. (2016) that the scale and nature of the data should match model complexity.

Turning to the area of link prediction, it is relatively underexplored in the energy domain. The majority of studies in this realm have paid attention to bibliographic, biological, and social networks; for example, deep-learning-based metrics were introduced by Nowell & Kleinberg (2007) and Zhang et al. (2016) for predicting future links in co-authorship and citation networks. However, the methodology in these papers doesn't include macroeconomic, geopolitical, or semantic features and is totally structural.

Guan et al. (2016) employ a link prediction model to investigate potential trade connections in the global crude oil market using the framework of complex network theory. They aim to reveal hidden relationships between nations that, despite not presently participating in crude oil trade, have a significant probability of forming trade partnerships in the future. The authors emphasized that although there is widespread employment of link prediction in different networks, such as social, biological, and transportation, its potential in examining potential trade connections within the crude oil trade network remains substantially unexplored. A significant ingenious feature of their approach is the incorporation of the count of shared trade partners among nations as a determinant for potential trade relationships. While Guan's paper focuses on the number of common trade partners as the major linking motivation behind crude oil trade connections, this research emphasizes non-structural variables as features for predicting active and non-active trade links.

By including international trade networks, Silva et al. (2024) demonstrated the use of ML for economic prediction. Similarly, the policy and institutional implications of AI-driven trade analytics have been aptly highlighted by Ozturk (2024). There are several studies that employ one or another machine learning method for link prediction. A groundbreaking methodology was employed by Wohlfarth & Ichise (2008) in order to predict connections within social networks. The authors utilized semantic and event-based parameters. This strategy is exemplary in the sense that they refrained from relying solely on network-based variables or basic node attributes but instead integrated them with semantic and temporal events. This approach enhanced the predictive framework. By employing a supervised machine learning approach, particularly decision trees, they tried to predict potential linkages within the social

network under consideration, empowering the model to uncover intricate and non-linear patterns in the data that influence connections.

A novel future link prediction strategy was employed by Han et al. (2015) for Online Social Networks (OSNs) involving new members in the social network who are yet to create links. The rigorous process entails procuring data on social characteristics using the limited data accessible from new users and the comprehensive data available from current users, and then employing a Support Vector Machine (SVM) to train a predictive model for linkages. Tuninetti et al. (2017) utilize a gravity-law model that integrates various country-specific characteristics (such as population, GDP, and agricultural efficiency) and link-specific attributes (e.g., geographical distances) to predict trade links in a virtual water trade network (VWTN).

In contrast to numerous network models that necessitate a comprehension of the network's structure to predict link establishment or termination, the approach of Tuninetti et al. (2017) can predict both the presence and strength of links based solely on external country and link attributes, which is especially advantageous for analyzing dynamic networks that undergo topology modifications. Tuninetti's approach to predicting trade links is similar to this paper in the sense that both approaches do not require prior information about the network's topology; however, they differ in the empirical models employed for prediction purposes.

3. Data

The dataset incorporates crude oil trade linkages from 2002 to 2021 of 12 major oil exporting countries commanding a total of 87.5% of global oil reserves as per BP Statistical Review of World Energy (2021). The crude oil exporters included in the study are - Algeria, Canada, Iran, Iraq, Kuwait, Libya, Mexico, Saudi Arabia, Russia, UAE, USA, and Venezuela with their respective reserves' percentages in 2021 being 0.7%, 9.7%, 9.1%, 8.4%, 5.9%, 2.8%,

Table 1: Variables and their corresponding data sources

S.No	Variables	Data Sources
1	Crude oil export value	UNcomtrade, Trademap
2	Past crude oil export value	UNcomtrade, Trademap
3	Number of treaties	United Nations Treaty Collection (UNTC)
4	Average tone (Avgtone)	Global Database of Events, Language, and Tone (GDELT)
5	Goldstein Scale	Global Database of Events, Language, and Tone (GDELT)
6	Sanctions	Global Sanctions Database (GSDB)
7	GDP importing country	World Bank
8	GDP exporting country	World Bank
9	Oil reserves exporting country	British Petroleum (BP), Organization of the Petroleum Exporting Countries (OPEC)
10	Oil reserves importing country	British Petroleum (BP), Organization of the Petroleum Exporting Countries (OPEC)
11	Sea distance	Centre d'Études et de Recherches sur le Développement International (CERDI)

0.4%, 17.2%, 6.2%, 5.6%, 4.0%, and 17.5%. The dependent variable is trade link formation, which is a categorical variable taking the value of 1 when the crude oil trade value is positive between an exporting and importing country pair in a given year. Alternatively, it assumes value 0 for those country pairs that traded in the immediately preceding year and are not trading in the current year. Table 1 shows the various variables used in the dataset and their respective sources.

The data on crude oil trading (HS code 270900) has been taken from UN Comtrade and Trademap databases and captures active and non-active trading links of the 12 crude oil exporting countries for the concerned period. HS code refers to the Harmonized System code and the no. 270900 pertains to petroleum oils and oils derived from bituminous minerals in their crude form.

The number of treaties built up temporally and cumulatively between different crude oil trading country pairs has been derived from the United Nations Treaty Collection (UNTC) database. It includes a broad array of agreements between various trading partners and is not constrained to trade-related pacts. The nexus of crude oil trade connectivities among importing and exporting countries is built on economic, political, military, financial, bilateral, and multilateral dimensions, which are diverse and have far-reaching impacts. The presence of an extensive number of treaties between country pairs signifies mutual understanding and cohesive relationships. This aspect captures relations between different countries globally more holistically as compared to the sole focus on trade-related treaties. For instance, military-based agreements could cultivate inexpungable trade relations and financial treaties could lead to improvements in financial capabilities, thus further bolstering trading prospects.

The Goldstein scale and AvgTone parameters corresponding to different trading country pairs and different years have been extracted using GDELT 1.0 Event Database from the Global Database of Events, Language, and Tone (GDELT) using country pairs as actor codes (Leetaru, 2013). In the realm of computational analysis of worldwide news data, the GDELT is considered a foundational repository and provides a thorough perspective through which the emotional tone of media coverage can be understood. The AvgTone (Average Tone) parameter captures the sentiment of a news article associated with a particular event or topic obtained by employing Natural Language Processing (NLP) techniques, which identify words carrying positive or negative sentiments based on a sentiment lexicon.

The range of the AvgTone parameter is from -100 to +100, with negative values denoting negative sentiments and positive values denoting positive sentiments. This study incorporates the AvgTone parameter as the annual average value of average tone corresponding to particular country pairs as actor codes and based on the events occurring between them for that particular year. The Goldstein scale designates numerical values to different categories of events based on their expected repercussion on the geopolitical arena, with magnitudes lying from -10 to +10, where negative values indicate actions expected to escalate conflict and instability, while positive values are related to the promotion of peace and stability. It functions within the framework of the Conflict and Mediation Event Observations (CAMEO) coding system, which already has an inherent score linked to a particular category of event having a specific CAMEO code; hence, the Goldstein scale score is not based on any computational analysis done by GDELT itself.

The CAMEO code classification is a comprehensive classification system categorizing all kinds of events with a geopolitical consideration. Country pairs engaging in diplomatic cooperation are assigned a positive score, and those engaging in an assault are assigned a negative score. The present research extracts the average annual Goldstein scale value from the GDELT dataset corresponding to different country pairs as actors and different trading years. A seminal study by Voukelatou et al. (2020) with reference to the calculation of the Global Peace Index (GPI) employed data from the GDELT database and identified AvgTone and Goldstein scale parameters as effective components in the construction of GPI. Their study connotes prominent progress as the conventional data sources have frequent delays and high acquisition costs.

Sanctions represent a categorical variable derived from the Global Sanctions Database (GSDB) (Syropoulos et. al., 2023) denoting a value of 1 when mutual or unilateral sanctions are imposed by exporting or importing countries, otherwise recorded as 0. These sanctions can be multilateral, bilateral, or unilateral, with only one condition that the particular crude oil trading country pair must be present either as a sanctioned country or a sanctioning country. Since crude oil is a commodity interlinked with geopolitical dimensions, the imposition and removal of sanctions is a crucial factor in the formation and dissolution of trade links between trading country pairs. Irrespective of their main motivation, sanctions have prominent and wide-ranging effects on political and economic considerations globally.

A pioneering work by Fischhendler et. al. (2017) analyzed energy sanctions between 1973 and 2002 and concluded that the imposition of sanctions typically targets the distribution segment of the energy supply chain and is often paired with non-energy sanctions. By incorporating a myriad of sanctions, including economic, trade, diplomatic, military, financial, and others, this study elaborately evaluates the bilateral relations between the trading country pairs.

The famed gravity equation of international trade says that the volume of trade flow between two countries is directly proportional to the product of their GDPs and inversely proportional to the square of the distance between them. To this end, the present research integrates the GDPs of the exporting and importing nations as node-specific features obtained from the World Bank.

Given that a significant portion of global crude oil transactions transpire via maritime pathways, this study has included the sea distances separating various pairs of trading nations as a metric for distance. The Centre d'Etudes et de Recherches sur le D´eveloppement International (CERDI) computes the distance across seas between nations by utilizing the most direct maritime pathways linking the primary ports of the countries in question. This strategy typically entails the identification of the key ports (or those with the highest trade volume on an international scale) of each nation, followed by the computation of the shortest viable distance over water connecting them (Bertoli et. al., 2016). The methodology involves using the geographic coordinates of the identified ports to provide a foundation for the measurement of distance. CERDI employs mathematical models to calculate the sea distance between countries, keeping in mind feasibility, borders, and other hindrances.

This study has incorporated oil reserves data from OPEC (OPEC, 2023) and BP (Energy Institute, 2023) for both exporting and importing countries as node-based characteristics. Oil reserves are a core indicator of a nation's energy security and its capacity to engage in the

global oil market in the long run. In the case of countries that export oil, the magnitude of reserves not only demonstrates their ability to maintain exports but also impacts their standing and authority in the international oil trading systems. Conversely, the reserves of countries that import oil reveal their susceptibility to supply disruptions and their consequent requirement to diversify trade connections. Oil-exporting nations with large reserves are motivated to establish trade relations corroborating their economic contentions and geopolitical influences. In contrast, substantial reserves for importing countries can serve as a strategic buffer, thus reducing reliance on external sources and potentially reshaping their trade network dynamics.

Finally, the past crude oil export value of a particular trading country pair extracted from UNComtrade and Trademap databases represents the value of crude oil traded in the immediate previous year. Recent crude oil trade statistics among nations tend to provide an unequivocal indication of the existing trade dynamics and interdependencies, and in turn depict the most current and relevant economic, political, and logistical aspects affecting trade.

4. Methodology

This research employs six distinct classification-based machine learning and deep learning models – Logistic Regression, Logistic Ridge Regression, Random Forest Classifier, SVM, CNN, and ANN. This study purposefully employs this diverse set of algorithms in order to grasp and compare varied dimensions of crude oil trade linkages. Logistic Regression serves as a baseline model that imposes a linear decision boundary that is quite handy in comprehending first-order relationships between features. In order to mitigate the risk of multicollinearity among the independent variables, Logistic Ridge Regression is particularly relevant.

The Random Forest model has been selected for its capability to capture nonlinear intricacies and robustness to noise and outliers, given the complex nature of geopolitics and the corresponding uneven influence on trade linkage dynamics. The model usually provides a resilient learning mechanism, particularly for structured data, and has the ability to generalize well. The strength of SVM lies in handling high-dimensional, sparse spaces and its robust capability to delineate patterns in the presence of subtle non-linear separability. The rationale behind the implementation of DL models is their ability to model complex nonlinear relationships and structural dependencies in the case of structured datasets. Further, in order to study the dynamic behavior of crude oil trade linkages comprehensively, this research uses four network-based measures, which include link turnover rate, substitution index, reactivation rate, and persistence ratio.

Various classification-based model evaluation metrics have been used to assess the effectiveness and robustness of the employed models, which cater to different aspects. The metrics are Accuracy, F1 Score (harmonic mean of precision and recall), and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These are pivotal for validating and comparing the predictive capabilities of all the models from different statistical perspectives[1]. The methodology in this study employs two distinct processes involving initial training on 80% of the dataset (with train-test split respecting the temporal order) and testing on the remaining 20%, which is the baseline result, and subsequent meticulous validation

[1] For a more detailed discussion, refer to the appendix.

using a series of rolling forward or walking forward time-series cross-validation cycles. The aim is to optimize both the performance and generalizability of the models.

Keeping in mind the temporal nature of the trade linkages under consideration, traditional cross-validation methods are not suitable as they could unintentionally produce look-ahead bias, a scenario in which the model is trained on the future year data points but is tested on the past year data points. This is a consequence of the mixing of data points across all periods. To properly validate and assess the robustness of the classification models, rolling-forward time series cross-validation has been employed for different train-test cycles. Four distinct variations of this method have been employed in this study, which include one-year train-test cycles, two-year train-test cycles, three-year train-test cycles, and four-year train-test cycles.

With an aim to better grasp the temporal unraveling of crude oil trade linkages, this study employs a set of network-based measures to capture the year-to-year trade dynamics in greater depth. The main motivation behind these measures is to provide insights into the strategic behavior, stability, and volatility of oil-exporting countries within the global energy trade network. Incorporating these measures enables us to analyze how the links evolve over time, how exporters adapt, and helps in capturing the variability of trade stability across countries and over time. To calculate these metrics, each dyad in a particular year has been tagged with one of the five dynamic statuses: formed, dissolved, re-established, active, and inactive.

A dyad appearing in a year t has been tagged as formed if it appeared only in year t and there was no prior activity. If a trade link present in the year t-1 has been broken in the year t, then this dyad has been tagged as dissolved in the year t. A trade link has been tagged as re-established if it reappears in the year t after being dissolved and remaining inactive in any year prior to t. For tagging a trade link as active, the link must remain stable and continue from year t-1 to t. Lastly, if a dyad was absent in the year t and was also absent in the year t-1, then it has been tagged as inactive[2] in year t. The following subsections further elaborate on the models and the network-based metrics employed in this study.

4.1. Logistic Regression

Logistic Regression is a machine learning model employed in cases when the dependent variable is binary. Its purpose is to predict the likelihood of link formation between a country pair based on multiple node-based and edge-based non-structural predictors. The outcome variable T_{ijt} denoting the existence of a trade link is binary and assumes value 1 when a trade link is present between the exporting country i and importing country j at time t, otherwise, it is 0. So, in essence, Logistic Regression gives the probability that the link will be formed between a country pair at a given time. Technically, we can represent it using the following equation:

[2] The detailed temporal network tags can be seen for each oil-exporting country through trade status plots given in the appendix (Figures A.14 – A.25).

$$log\left(\frac{P(T_{ijt}=1)}{1-P(T_{ijt}=1)}\right) = \beta_0 + \beta_1.Treaties_{ijt} + \beta_2.V2_{ijt} + \beta_3.Gold_{ijt} + \beta_4.Sanc_{ijt} + \beta_5.GDP_{imp_{jt}} + \beta_6.GDP_{exp_{it}} + \beta_7.Res_{exp_{it}} + \beta_8.Res_{imp_{jt}} + \beta_9.Sead_{ij} + \beta_{10}.Pastexp_{ijt} + \beta_{11}.Year + \beta_{12}.Treaties_{ijt} * Sanc_{ijt} + \beta_{13}.V2_{ijt} * Gold_{ijt} + \beta_{14}.Sead_{ij} * Pastexp_{ijt} + \beta_{15}.Treaties_{ijt} * V2_{ijt} \quad (1)$$

Here, $log\left(\frac{P(T_{ijt}=1)}{1-P(T_{ijt}=1)}\right)$ denotes the log odds ratio of the probability of crude oil trade link formation, $P(T_{ijt}=1)$ denotes the probability of the existence of a crude oil trade link between countries i and j in year t, β_0 represents the intercept, and β_1, β_2,..., β_{15} are the coefficients that measure the impact of each predictor on the log odds of the trade link's existence. The predictor $Treaties_{ijt}$ captures the cumulative number of unilateral or multilateral treaties between exporting country i and importing country j at time t, $V2_{ijt}$ takes into account the average sentimental tone for all the events that occurred between exporting country i and importing country j at time t, $Gold_{ijt}$ shows the average Goldstein scale value of all the events that occurred between exporting country i and importing country j at time t, $Sanc_{ijt}$ is a categorical variable assuming value 1 when any kind of bilateral, unilateral, or multilateral sanction exists between exporting country i and importing country j at time t, otherwise it is 0, $GDP_{imp_{jt}}$ and $GDP_{exp_{it}}$ denote the corresponding GDPs of exporting country i at time t and importing country j at time t, respectively, $Res_{exp_{it}}$ and $Res_{imp_{jt}}$ capture the corresponding crude oil reserves of exporting country i at time t and importing country j at time t, respectively, $Sead_{ij}$ presents the sea distance between exporting country i and importing country j as calculated by CERDI, $Pastexp_{ijt}$ reflects the crude oil trade value exported by country i to country j at time t-1, and Year represents the time variable.

4.2 Logistic Ridge Regression

This research employs the Logistic Ridge Regression model as a robustness check and also to counter the issue of overfitting in the Logistic Regression model. It is also termed as L2 regularization or ridge regularization. Mathematically, the ridge regularization relates to minimizing the following loss function:

$$L(\beta)_{ridge} = -\frac{1}{N}\sum_{i=1}^{N}\left[T_{ijt}\,log(\hat{T}_{ijt}) + (1-T_{ijt})\,log(1-\hat{T}_{ijt})\right] + \lambda\sum_{k=1}^{p}\beta_k^2 \quad (2)$$

Here, $L(\beta)_{ridge}$ is the loss function associated with ridge regularization, λ is the regularization strength, p is the total number of predictors in the model, N is the number of observations, and β is the coefficient of the model[3]. The hyperparameter λ is chosen by employing grid search combined with time series cross-validation.

4.3 Random Forest Classifier

[3] For a more detailed discussion, refer to the appendix.

This research employs the Random Forest Model, which is a non-linear classification-based link prediction algorithm. The Random Forest Classifier is an ensemble of decision trees commonly trained with the 'bagging' method, which is used to reduce variance within a noisy dataset. For the final classification task, the prediction of cessation or formation of crude oil trade link is made by majority voting involving all the decision trees. This study has used grid search combined with time series cross-validation to optimize hyperparameters related to number of trees in the forest, maximum depth of each tree, minimum number of samples required to split each node, minimum number of samples required to be at a leaf node and complexity parameter used for minimal cost-complexity pruning (α) which essentially adds a penalty proportional to the number of terminal nodes. Further, the Gini index has been used as the splitting criterion[4].

4.4 Support Vector Machine (SVM)

This research employs another famous supervised machine learning model, termed as Support Vector Machine (SVM) classification model. This model runs on an optimization framework whose objective function is to find a hyperplane in the Euclidean space defined by the normal vector $\boldsymbol{W}$ and the bias b. This plane is situated in the Euclidean space in a way such that it separates the data points in the crude oil trade linkage dataset into two classes (trade link existence T_{ijt}=1 or absence T_{ijt}=0) with maximum margin. Mathematically, the optimization problem with constraints is written as:

$$\begin{gathered} \min_{\boldsymbol{W},b} \frac{1}{2}\boldsymbol{W}^T\boldsymbol{W} + C\Sigma\xi_{ijt} \\ \text{s,t. } T_{ijt}\left(\boldsymbol{W}^T\phi\left(X_{ijt}\right) + b\right) \geq 1 - \xi_{ijt}, \quad \forall i,j,t \\ \xi_{ijt} \geq 0 \end{gathered} \tag{3}$$

Here, $\phi\left(X_{ijt}\right)$ is a transformation function or kernel function that allows the projection of the features into a higher-dimensional space, thus facilitating the separation of classes with a linear boundary. ξ_{ijt} is a slack variable introduced to handle data misclassification and depicts that the margin of SVM is soft. C is a regularization parameter controlling the trade-off between maximizing the margin of SVM (more misclassifications) and minimizing the classification error. $\boldsymbol{W}$ is a vector involving the weights or coefficients defining the orientation of the hyperplane in the feature space, and b is the bias term that specifies the distance of the hyperplane from the origin[5]. The hyperparameters of this model have been tuned using grid search combined with time-series cross-validation.

4.5 Artificial Neural Network (ANN)

An ANN is the simplest deep learning model inspired by the operations of biological neurons and their information dissemination mechanism. It encompasses various layers of nodes, also known as neurons, that are interconnected and tasked with transmitting information through weighted connections. The model is typically organized into three types of layers involving the input layer, the hidden layer, and the output layer. The input layer comprises the input features of the dataset, hidden layers are intermediate layers responsible for transforming the

[4] For a more detailed discussion, refer to the appendix.
[5] For a more detailed discussion, refer to the appendix.

input data through a series of linear as well as non-linear operations, and the output layer is the final layer that produces an output that could be a single value (in case of regression) or a probability distribution over different classes (in case of classification)[6]. In this paper, the ANN model attempts to minimize the binary cross-entropy loss function given by equation (1) by adjusting weights and biases of the model using an optimization algorithm.

$$L(ANN) = -\frac{1}{N}\sum_{i=1}^{N}\left[T_{ijt}\, log\left(\hat{T}_{ijt}\right) + \left(1 - T_{ijt}\right) log\left(1 - \hat{T}_{ijt}\right)\right] \quad (4)$$

4.5 Convolutional Neural Network (CNN)

A CNN is a particular type of deep learning model mainly designed to learn spatial characteristics from the input data by employing various intricate operations. The core foundation of a CNN model is the convolutional layer, in which a set of kernels or filters slide over the input data in order to extract local features. The output of this convolutional layer is a feature map in which each element constitutes the result of employing the filter at different spatial locations. The resultant output after the convolutional operation is passed through an activation function, which introduces nonlinearity into the model. Further, pooling layers are employed to reduce the dimensionality and to preserve the most important information so that the computation is more efficient and the model is less sensitive to minor shifts and distortions in the dataset. Lastly, after several convolutional and pooling layers, fully connected layers are used, having connections with every activation in the previous layer. The output of the fully connected layers is fed into a SoftMax function for multi-class classification or a sigmoid function in the case of binary classification. The model is trained using backpropagation combined with gradient descent algorithms, in which parameters are updated upon minimizing the cross-entropy loss function[7].

4.6 Turnover Rate

The overall dynamism of an exporting country's trade network is quantified by the turnover rate. Greater volatility and a relatively more fluid network structure are shown by a higher turnover rate, whereas a lower value represents trade stability. Specifically, this metric aids in assessing whether a nation's oil trade relationships are persistent or frequently reconfigured. For an exporting country i and year t, the turnover rate can be mathematically expressed as:

$$Turnover\ rate_{i,t} = \frac{Number\ of\ formed\ links_{i,t} + Number\ of\ dissolved\ links_{i,t}}{Number\ of\ active\ links_{i,t}} \quad (5)$$

4.7 Substitution Index

An exporting country's propensity to replace dissolved trade links with new ones is captured by the Substitution index. An aggressive substitution strategy is represented by a substitution index value greater than 1, as exporters actively search for new markets to offset previous links. This metric comes particularly handy in comprehending the response of countries to geopolitical disruptions and external shocks. For an oil-exporting country i and year t, it can be written mathematically as:

[6] For a more detailed discussion, refer to the appendix.
[7] For a more detailed discussion, refer to the appendix.

$$Substitution\ Index_{i,t} = \frac{Number\ of\ formed\ links_{i,t}}{Number\ of\ dissolved\ links_{i,t}} \quad (6)$$

4.8 Reactivation Rate

The frequency with which dissolved trade links are re-established is being captured by the reactivation rate, reflecting the resilience and re-engagement behavior of exporters. Particularly, it demonstrates the fact that how often the oil-exporting countries return to past trade partners after disruptions due to logistical, economic, or political barriers. Mathematically, the reactivation rate for an exporting country i and year t can be expressed as:

$$Reactivation\ Rate_{i,t} = \frac{Number\ of\ re-established\ links_{i,t}}{Number\ of\ previously\ dissolved\ links_{i,t-1}} \quad (7)$$

4.9 Persistence Ratio

The percentage of trade links retained from the previous year relative to the total number of dyads is indicated by the persistence ratio, serving as a direct measure of assessing the continuity and stability of an exporter's trade relationships. High persistence ratios are a benchmark of long-standing export ties and often reflect stable geopolitical alignments or contractual obligations depicting long-term relationships. On the other hand, opportunistic or strategic dynamism is demonstrated by a low persistence ratio. For an exporting country i and year t, this metric can be written mathematically as:

$$Persistence\ Ratio_{i,t} = \frac{Number\ of\ active\ links_{i,t}}{Total\ dyads_{i,t}} \quad (8)$$

5. Results and discussions

5.1. Baseline results

Table 2: Classification Report of Logistic Regression

	Precision	Recall	F1 score
Class 0	0.75	0.95	0.84
Class 1	0.90	0.58	0.71
Accuracy			0.79
Macro avg	0.83	0.77	0.77
Weighted avg	0.82	0.79	0.78

Table 3: Classification Report of Logistic Ridge Regression

	Precision	Recall	F1 score
Class 0	0.76	0.95	0.85
Class 1	0.91	0.61	0.73
Accuracy			0.81
Macro avg	0.84	0.78	0.79
Weighted avg	0.83	0.81	0.80

Table 4: Classification Report of Random Forest

	Precision	Recall	F1 score
Class 0	0.87	0.86	0.87
Class 1	0.82	0.83	0.83
Accuracy			0.85
Macro avg	0.85	0.85	0.85
Weighted avg	0.85	0.85	0.85

Table 5: Classification Report of SVM

	Precision	Recall	F1 score
Class 0	0.73	0.98	0.84
Class 1	0.95	0.53	0.68
Accuracy			0.78
Macro avg	0.84	0.75	0.76
Weighted avg	0.83	0.78	0.77

Table 6: Classification Report of ANN

	Precision	Recall	F1 score
Class 0	0.77	0.91	0.84
Class 1	0.84	0.65	0.73
Accuracy			0.80
Macro avg	0.81	0.78	0.78
Weighted avg	0.80	0.80	0.79

Table 7: Classification Report of CNN

	Precision	Recall	F1 score
Class 0	0.80	0.83	0.81
Class 1	0.76	0.73	0.74
Accuracy			0.78
Macro avg	0.78	0.78	0.78
Weighted avg	0.78	0.78	0.78

Tables 2-7 represent the classification reports of all six models as a baseline result when 80% of the data in temporal order is used to train the models. Upon observing the results, it is clearly evident that each of the classification algorithms portrays distinct trade-offs between precision and recall across both classes, i.e., active and non-active trade links. A larger proportion of actual positives are effectively identified by the Logistic Regression model, but it is at the cost of increased false positives, which can be seen by higher recall over precision for class 0.

The model scores a higher accuracy of 90% in identifying predicted positives, which in turn signifies a focus on precision for class 1, but with a lower recall. On average, the model is correct 79% of the time with an average F1 score of 77%. Relatively, the Logistic Ridge Regression strikes a better balance between precision and recall across both categories, which is reflected in its improved accuracy of 81% and F1 score of 79%, suggesting that regularization is indeed playing a crucial role. Achieving an accuracy of 85% and an F1 score

of 85%, the Random Forest model is the best-performing algorithm in every metric as a result of its robustness in identifying both linear and non-linear intricate patterns aptly.

The model achieves the most balanced trade-off among all four algorithms employed in this study, although it prioritizes recall in class 0 and precision in class 1. On the other hand, while the SVM shows the highest recall in class 0 across all the models, it sacrifices precision. Turning to class 1, the model excels in precision, but it is at the cost of recall, suggesting that the model is quite reliable when it comes to predicting trade link formation, but its ability to identify all trade links is limited. On average, the model fares slightly poorer than the Logistic Regression variants, suggesting that the decision boundary is complex. Overall, since the objective of this research is to have balanced prediction across both categories, the Random Forest Classifier turns out to be the most reliable and robust model.

Comparing both the DL models, they show reasonably balanced classification performance as far as accuracy is concerned, with the corresponding values being 80% and 78%, respectively. ANN model has the ability to correctly identify a majority of non-active trade links, as indicated by its recall of 0.91 for class category 0; however, it comes at a tradeoff of reduced recall of 0.65 for active trade link category (class 1). But when the model does predict a trade link, it is usually correct, as can be seen from the precision of 0.84 in class 1, thus making it a reliable model as far as minimizing false positives for class 1 is concerned. Contrastingly, the CNN model demonstrates more balanced precision and recall across both class categories with a slightly lower accuracy of 78% relative to ANN. Achieving a precision of 0.80 and a recall of 0.83 in class 0 and a precision of 0.76 and a recall of 0.73 for class 1, the model leads to equal macro and weighted F1 scores across both classes, indicating a moderate capacity to differentiate between both classes fairly evenly.

The fact that ANN demonstrates higher accuracy than CNN relates to the bias of ANN towards the majority class. While the ANN model learns global patterns quickly, it tends to memorize dominant patterns in the case of limited data and imbalanced labels. Further, given that CNN treats each input feature as a spatial or channel dimension and has the appreciable capability to catch local correlation between features, it generalizes better across classes and gives a more balanced result.

But despite the architectural sophistication of these models, both of them underperform relative to the Random Forest Classifier employed in the author's earlier work in terms of all the evaluation metrics. Some of the possible reasons attributed to this underperformance can be related to the fact that DL models generally work best on large-scale datasets. Additionally, the limited temporal depth of the dataset due to data constraints, the aptness of the Random Forest and other traditional ML models in case of hand-crafted features with domain-specific knowledge, the increased risk of overfitting in complicated models, and optimization challenges have probably taken a toll on the performance of DL models.

Fig. 1 represents the AUCs of all the models for the case when 80% of the dataset is used for model training. The AUCs of Logistic Regression, Logistic Ridge Regression, Random Forest Classifier, SVM, CNN, and ANN models are 0.83, 0.84, 0.91, 0.85, 0.84, and 0.84, respectively. It is evident from these magnitudes that the Random Forest Classifier again shows the best performance and has the highest discriminatory power in terms of both categories for the prediction of active and non-active trade links. This could be credited to the ensemble approach of the model, allowing it to capture complex nonlinear patterns present in

the dataset. Also, the SVM model performs better than both variants of Logistic Regression. This could be credited to the ensemble approach of the model, allowing it to capture complex nonlinear patterns present in the dataset. Also, the SVM model performs better than both variants of Logistic Regression.

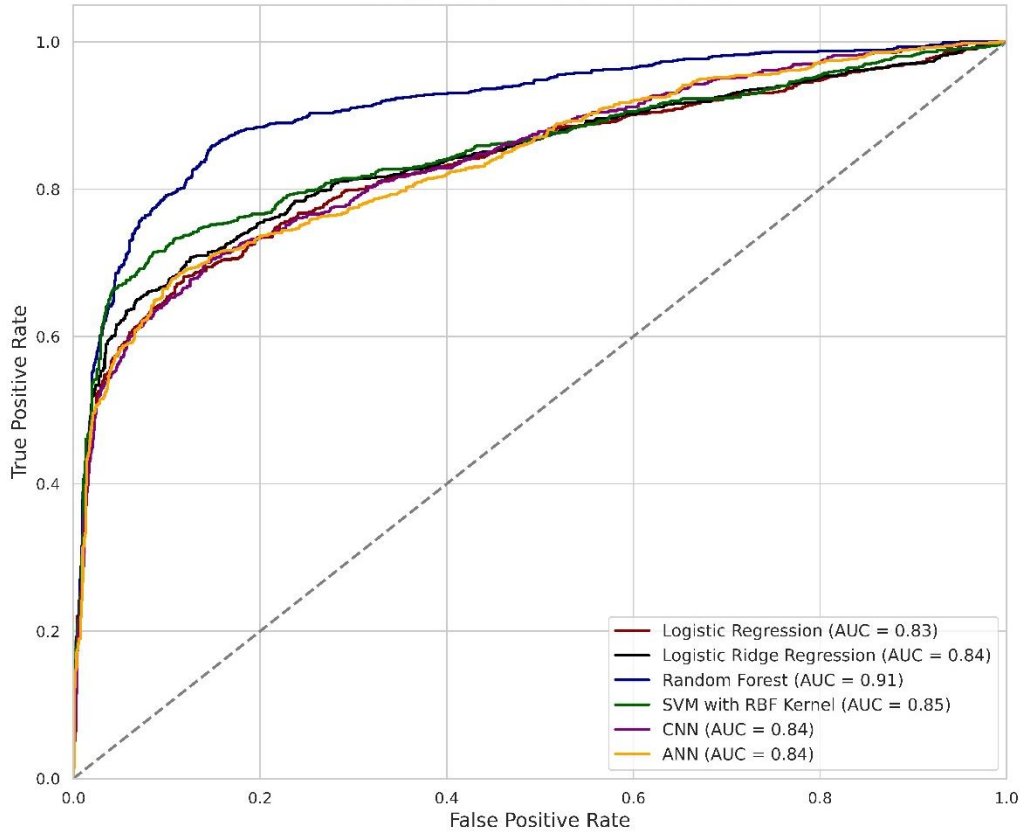


Fig. 1. AUC-ROCs of all the models when 80% of the dataset is used as the training set

ANN and CNN have a reasonably strong discriminatory power of 0.84. The performance of the ANN flattens out at higher thresholds, while it remains relatively stable in the case of CNN, depicting a better balance between sensitivity and specificity in the case of the latter model. This trend aligns with the presence of more balance between the precision and recall metrics of the CNN model observed earlier. But both models perform worse than the Random Forest and SVM models.

It is of paramount importance to illuminate the crucial aspect that although at first sight, the difference across different models may appear to be minimal, the models excel in different areas. For instance, given the high precision of the SVM model in predicting trade link formation, it is more reliable in confirming new links in spite of being conservative. Conversely, when it comes to identifying vulnerable trade partnerships, the Random Forest model offers better recall. Therefore, close performances across models reinforce confidence in the predictive findings, which strengthens the generalizability and resilience of this research through variations in decision boundaries, sensitivity to hyperparameters, and inductive biases.

5.2. Robustness of the models

The task that is of utmost priority in this study is of generalizability, applicability, and predictive power of the trained models to unseen data. The trained model must be consistent and less volatile in terms of predictive power. To test the robustness of all six models with respect to volatile geopolitics and temporal fluctuations, this research has utilized the power of a rolling forward time series cross-validation strategy. Figures 2, 3, and 4 show the test set metrics of all the models for a one-year train and one-year test cycle.

Both Random Forest and Logistic Ridge Regression tend to have higher accuracy, while the former is superior in terms of achieving higher peaks and consistency. The SVM model has the lowest accuracy, while all three models, i.e., Logistic Regression, Logistic Ridge Regression, and SVM, are highly volatile as compared to the Random Forest model. This is true across all three metrics except AUC, in which case, the Logistic Regression is the worst-performing model. The dismal performance of the SVM model may possibly be attributed to the fact that the decision boundary is too complex or is influenced by outliers. Both Random Forest and Logistic Ridge Regression models generalize well in the presence of noise.

Noticeable performance fluctuations across various test years can be seen for both DL models, thus exhibiting relatively lower stability. The performance of CNN in terms of relatively better F1 balance suggests that it is more resilient to short-term disruptions. While ANN attains almost comparable accuracy post-2015, it shows pronounced instability in earlier years, in turn, depicting higher sensitivity to shorter training windows. For two-year train and one-year test cycles, CNN maintains a slightly greater accuracy in most test years relative to ANN, particularly from 2017 onwards. A more consistent F1 trend is seen in CNN, given that this metric fluctuates sharply in early years for ANN, thus indicating instability in capturing class balance.

Notably, all the models struggled in the years 2003, 2008-09, 2020-21, and 2014-15, which shows that these years were influenced by significant geopolitical developments and are different from the trend of the entire dataset[8]. Superior consistency over volatile geopolitical periods is being demonstrated by the Random Forest Model, whereas other models show more volatility. Moreover, the stability across models is actually a strength and not a limitation, as it signifies the crucial fact that the underlying signal is strong enough to be detected across different methodological lenses.

The results presented in the main body of this paper focus only on a representative set of rolling forward train-test cycles, i.e., one-year train and one-year test cycles, in the interest of brevity[9]. Relative to one-year train and one-year test cycles, the performance of all the models has improved for longer training windows. Moreover, these plots reveal that the Random Forest model has a greater adaptability and learning capacity even in the case of cyclical patterns involving a diverse range of events.

[8] This can further be observed by referring to the average scores of all the models across all test years shown in the appendix (Table A.1).

[9] Comprehensive results for all other rolling forward train-test cycles, along with corresponding tables and figures, are included in the appendix for further reference (Figures A.5 – A.13, Tables A.2 – A.4).

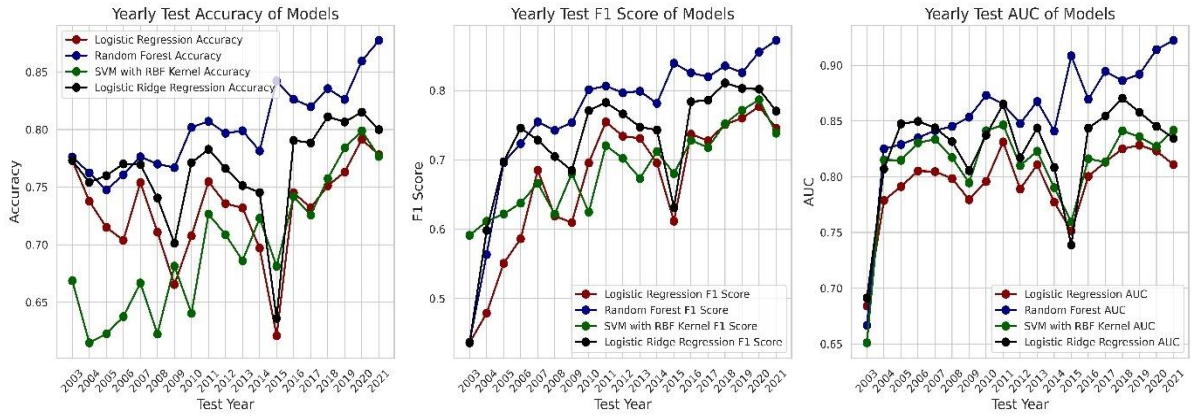


Fig. 2. Accuracy, F1 score, and AUC plots for Logistic Regression, Logistic Ridge Regression, Random Forest classifier, and SVM for one-year train and one-year test cycles

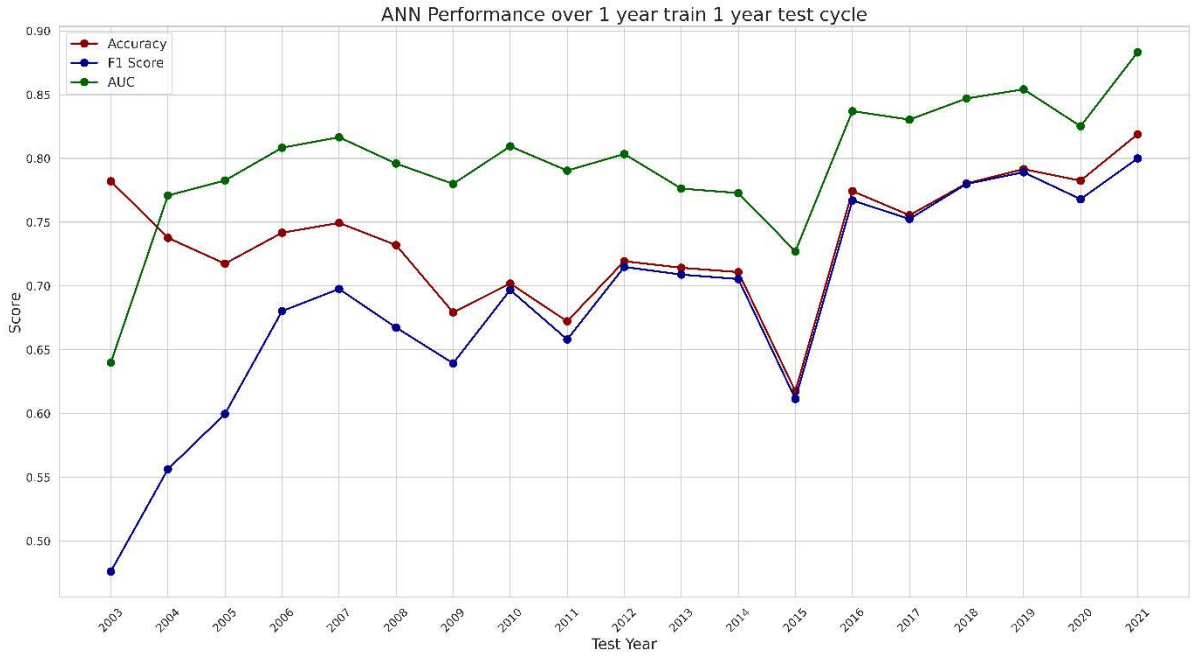


Fig. 3. Accuracy, F1 score, and AUC plots for ANN for one-year train and one-year test cycles

The three-year train and one-year test window[10] offers a broader learning context in terms of catching persistent geopolitical and economic shifts. While ANN achieves higher peak accuracy, CNN exhibits greater consistency, although the peak value is marginally lower relative to ANN. This is true for the F1 score, too, as CNN is more balanced with fewer dips, given the fact that ANN delivers impressive F1 scores in 2010-11 and 2019-21, but with overall higher volatility. In terms of threshold discrimination and overall model confidence, CNN again outperforms ANN, as can be inferred by AUC values. Overall, as the length of the temporal window is increased, both DL models show improvement in their performances; however, they have their own nuanced strengths and limitations.

[10] Refer to the appendix for the plot.

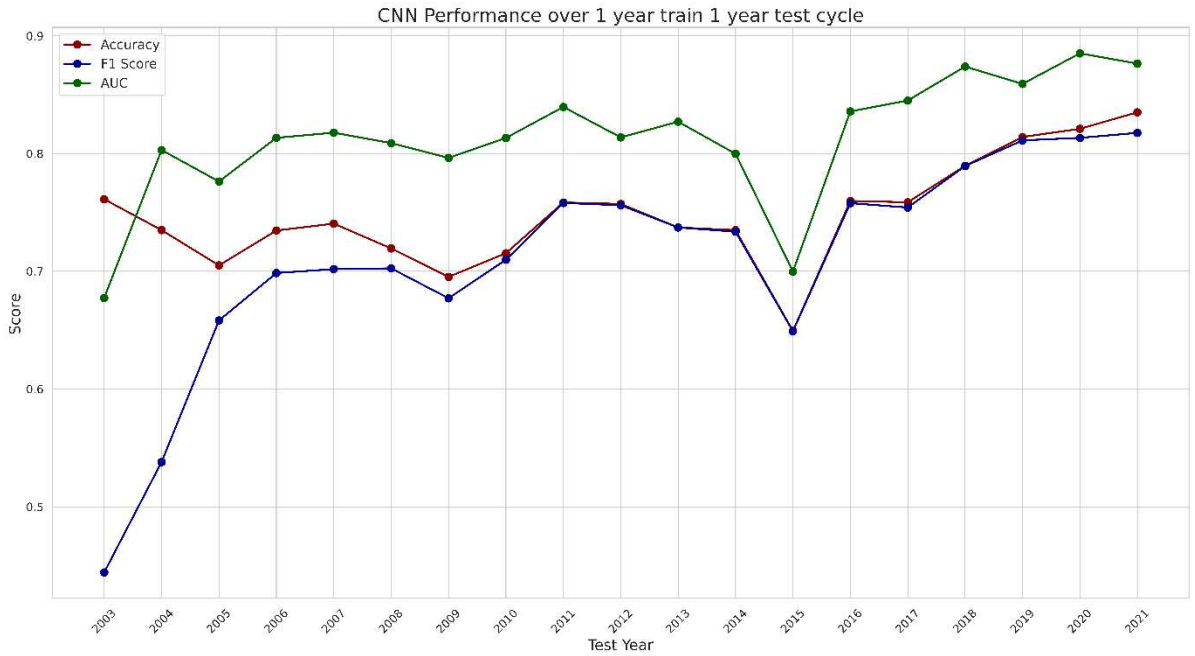


Fig. 4. Accuracy, F1 score, and AUC plots for CNN for one-year train and one-year test cycles

ANN has the ability to capture global non-linear feature interactions, due to which it excels in stable and predictable geopolitical environments, which is especially true in recent years (2016-2021). But the model struggles in the case of abrupt geopolitical shifts. This trend can also be observed through the increased imbalance between sensitivity and specificity for turbulent years.

Whereas, in this tabular-spatiotemporal context, the CNN model proves to be remarkably effective. The strength of the model lies in its proficiency in grasping local feature dependencies and hierarchical structures, rendering it robust to volatility and apt at generalization across time. Consequently, the performance is smoother across years with decreased inter-annual fluctuation, specifically, during turbulent years like 2008-09, 2015-15, and 2020.

If we compare the results with traditional ML models, the Random Forest model clearly exhibits superior performance both in terms of year-on-year stability and overall predictive accuracy, and provides the best balance between precision and recall. This trend could be attributed to the algorithm's robustness to noise and outliers, its ability to model nonlinearities without extensive hyperparameter tuning, and its ensemble learning structure. Therefore, in economic domains characterized by moderately sized structured data and the need for robustness in the face of missing data or policy shocks, the Random Forest model should be the first choice.

No matter how robust a model is, it is imperative to understand that the performance of the models is also influenced and, in turn, determined by the volatile and unpredictable global events affecting trade linkage dynamics of crude oil, which in turn results in certain years standing out in terms of predictive difficulty. A dip in AUC and F1 scores is seen in the year 2003, possibly indicating that during this period, there were a lot of unpredictable geopolitical events. The years 2001, 2002, and 2003 are particularly marked by severe

geopolitical turmoils, with the 9/11 terror attack in the U.S., the Venezuelan oil production strike, and Iraq's invasion by the U.S. being the most prominent among all.

Patidar et al. (2024) analyzed these events in great detail. In the aftermath of the 9/11 attacks in the U.S. claimed by al-Qaeda, the U.S. launched a war against terrorism with the invasion of Afghanistan in October 2001, aimed at dismantling al-Qaeda. While Afghanistan is not a prominent oil-producing country, it is situated near Central Asia and thus contributes to the political instability of the region. The Venezuelan oil production strike between December 2002 and January 2003, owing to political conflicts, led to a drastic drop in the production of crude oil and had a sweeping impact on oil prices.

Also, the upsurge leading up to the Iraq war in 2003 was related to the war on terrorism launched by the U.S. globally. Iraq, being a significant oil producer prior to the war, suffered huge repercussions with large-scale disruption in crude oil production and thus supply. The intense risk and volatility induced by the war on terrorism had a direct impact on oil prices. With increased concerns about the security of critical infrastructure involving oil pipelines and shipping routes, many nations reevaluated their energy policies with an increased focus on diversification of their energy sources. Various train and test cycles show that these events had a long-term effect on crude oil link formation and dissolution.

One other particular scenario is the significant dip in performance around 2008, which can be attributed to the impact of the global financial crisis affecting global trade in general. This has been well explained by Hamilton (2009). He corroborates that the increase in oil prices during 2007 and in the early quarters of 2008 stemmed from the fact that there was an increase in demand coupled with stagnant production levels. Although it is noteworthy that, in contrast to the previous disruptions in the crude oil market, this time there was an absence of prominent decreases in oil supply, rather, production was stagnant. The coupling of supply-demand dynamics led to a sudden and sharp increase in demand, particularly in developing economies like China, resulting in record-high oil prices by the middle of 2008.

Hamilton (2009) further emphasized that this situation took a radical turn owing to the emergence of the global financial crisis in the latter part of 2008. The crisis triggered economic depression, reducing the oil demand globally, and when coupled with the speculative bubble, it further resulted in a sudden drop in oil prices. Explaining the repercussions, the author mentions that it prompted a transition towards sustainable energy resources and incentivized a reassessment of the energy security landscape. Moreover, this crisis led to a shift towards a more interconnected and interdependent oil market in which developing economies held a pivotal role in shaping the global geopolitical arena and, in turn, presented novel challenges for oil-exporting countries in managing their resources effectively amidst a volatile demand and supply scenario.

A noticeable dip in performance was further observed around 2011, which is in all likelihood related to the Arab Spring's impact on oil production and hence prices, owing to the increased geopolitical tensions in the major oil-producing regions of the world. As emphasized by Patidar et. al (2024), in 2011, the oil market was remarkably dictated by the Libyan crisis emerging within the context of the widespread Arab Spring movements traversing multiple Arab countries, coinciding with major oil-exporting regions of the world. About the crisis, Libya witnessed a drastic decline in its oil production, leading to an immediate burgeoning in oil prices, further amplified by concerns related to the availability of oil. The authors

explained the underlying cause and mentioned that the oil production of Libya predominantly catering to European markets, plummeted from approximately 1.6 million barrels per day to almost 0 at the peak of the intense conflict.

Also, a descent is perceived around 2014-15, which might be associated with the crash in oil prices due to excess supply and a slowdown in global demand, disturbing the demand and supply dynamics. This has been explored by Patidar et. al (2024) in depth. They attributed the decline in prices to multiple factors, including the rise in U.S. oil production on the grounds of efficiency enhancements in shale oil extraction, coupled with decreasing geopolitical tensions. Additionally, the approach of OPEC countries to increase production was aimed at capturing a larger market share and outcompeting the rivals, and it led to a further fall in crude oil prices.

The COVID-19 pandemic also impacted the performance of the models, which further demonstrates that these models might not have been able to correctly capture the nonlinear changes in trade dynamics during this period. This also signifies that the crisis possibly led to a shift in trade priorities in the sense that countries started focusing on essential goods. Also, some countries may have formed new trade alliances or dissolved existing ones in order to adapt to the changing global dynamics, which is hard to detect since the models are trained on the pre-pandemic data. On the contrary, Random Forest has done relatively better since it is a relatively flexible model and is more apt at capturing sophisticated nonlinear patterns.

These analyses demonstrate that crude oil trade linkages are not solely linear but are dictated by conditional, nested, and hierarchical relationships depicting structured complexity with local and global interactions. The volatility of models suggests that link formation is not purely rational or economically deterministic; it is path-dependent and disruption-sensitive. While crude oil trade patterns are learnable, they are not fully predictable as they don't follow static rules, thus affirming their intricate adaptive system driven by policy shifts, geopolitical alliances, and economic fundamentals. These nonlinear dynamics, structural breaks, and interdependent feature interactions are hard to capture by traditional econometric methods relying on various restrictive assumptions of stationarity, linearity, and parametric specifications.

5.3. Analysis of network-based metrics

Figure 5-8 demonstrates the year-wise link turnover rate, substitution index, reactivation rate, and persistence ratio, respectively, of various crude oil exporting countries[11]. Intricate geopolitical, policy-driven, and infrastructural transformations in global oil trade networks are being revealed by these metrics. The discovery of oil sands and pipeline expansions led to the high turnover rates for Canada in the early 2000s, while the disruptive repercussions of the US sanctions are explicitly visible through the trends shown by Iran. In contrast, low turnover and high persistence are being shown by countries like Saudi Arabia, Kuwait, and Russia, signifying the existence of long-term contracts, specifically with Asian importers.

If we look at the substitution patterns, it captures the demand-side volatility caused by geopolitical disruptions. On the one hand, Canada showed higher substitution in 2007 and 2021 due to environmental and policy shifts; on the other hand, during periods of sanctions,

[11] For a more detailed country wise interpretation of these metrics kindly refer to the appendix.

Iranian crude was repeatedly replaced by other exporters. Similarly, as a result of changing refinery preferences, declining production, and domestic unrest, both Libya and Mexico faced substitution peaks.

Episodic market re-engagements are captured by reactivation rates as observed clearly in Iraq post-2003, Iran following the Joint Comprehensive Plan of Action (JCPOA) in 2016, and the UAE post Arab Spring. But the durability of these reactivations was being compromised owing to geopolitical instability, such as in the case of Libya and Venezuela. The policy shift in Russia through the eastward diversification strategy as a result of strained EU relations can be observed by sustained reactivation rates post-2005. Focusing on the durability of trade ties, Algeria, Russia, and Saudi Arabia maintained durable trade relations as demonstrated through their high persistence ratios, in turn, depicting reliability and strategic positioning.

Comparatively, sanctions, political volatility, and domestic turmoil have resulted in a sharp decline in the persistence ratios of Venezuela and Iran. Moreover, both the US and Canada showcase asymmetric persistence owing to evolving export policies and a lack of market diversification, respectively. These trends increasingly reflect evolving energy strategies, external shocks, and policy reforms that reshape global crude oil trade architectures, which are vital for anticipating market realignments, vulnerabilities in the supply chain, and the geopolitical underpinnings driving the currents.

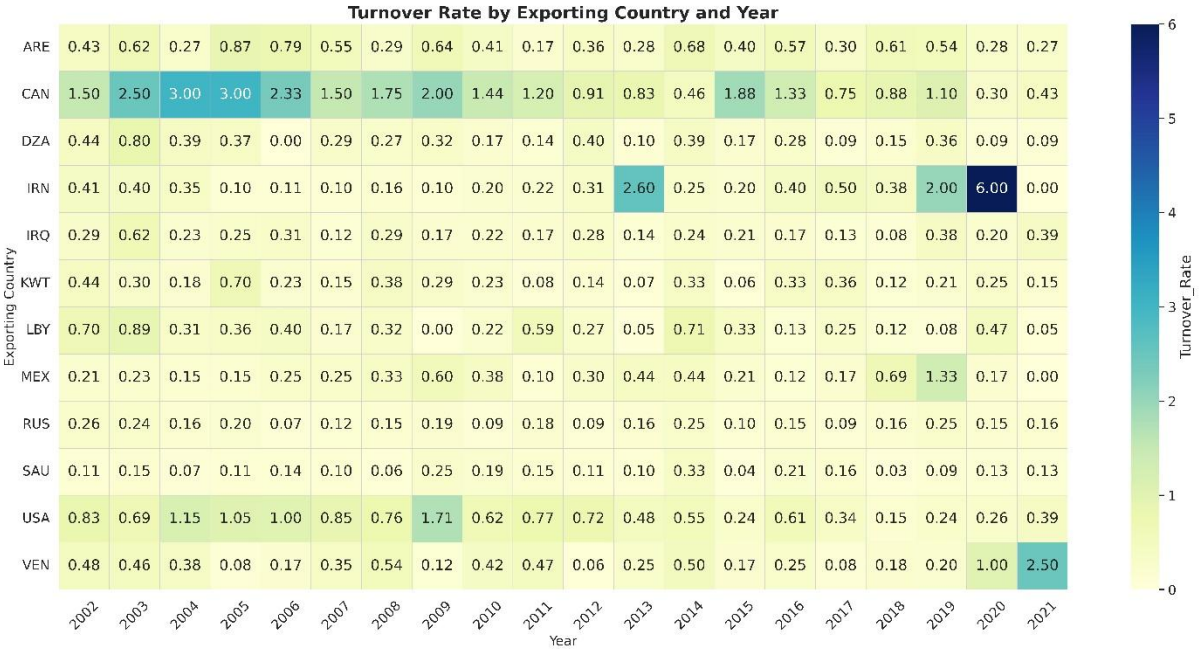


Fig. 5. Year-wise link turnover rates of crude oil exporting countries



Fig. 6. Year-wise link substitution index of crude oil exporting countries

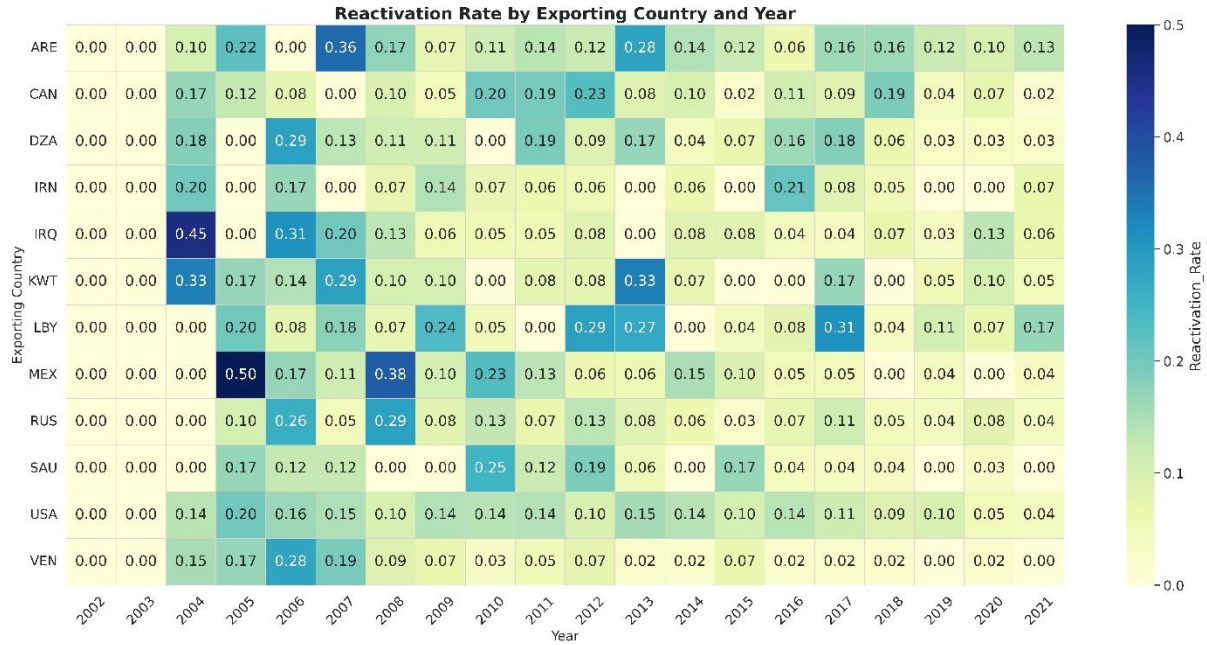


Fig. 7. Year-wise link reactivation rate of crude oil exporting countries

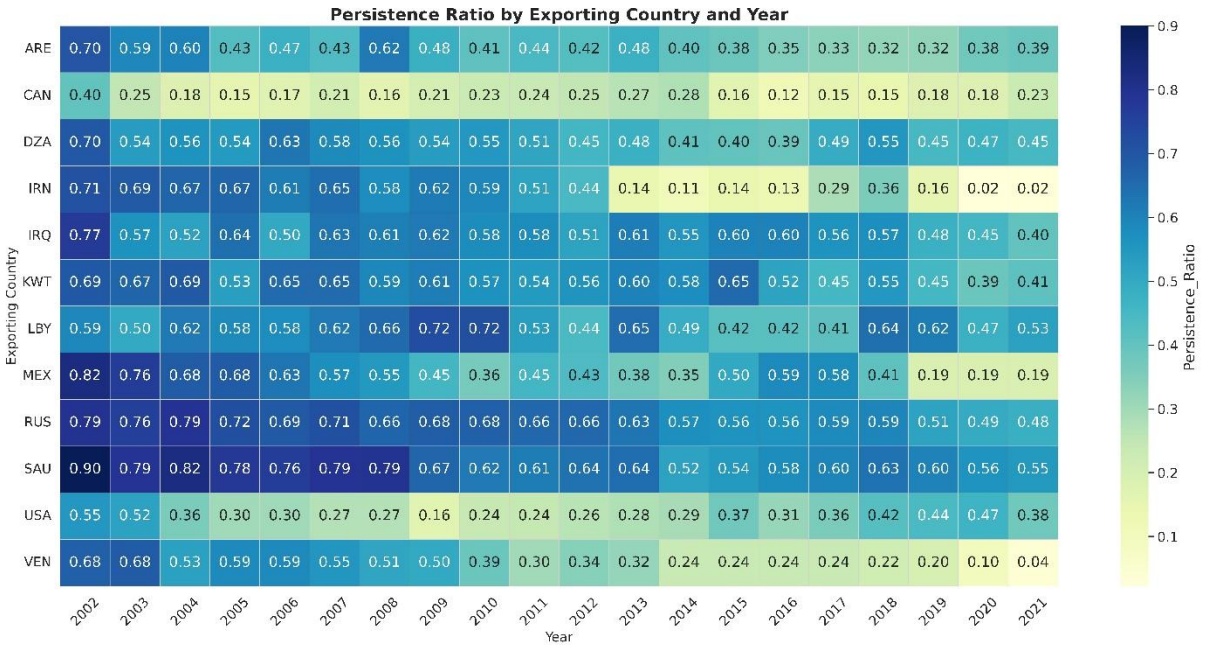


Fig. 8. Year-wise link persistence ratio of crude oil exporting countries

6. Conclusion and policy implications

This research enlarges the scope and applicability of machine learning and deep learning models to a crucial domain that is intricately intertwined with the nuances of geopolitical strategies, economic motivations, and energy security, and emphasizes that machine learning models are indeed capable of capturing the complexities and non-linearities well. Further, by employing a dataset spanning two decades and incorporating diverse economic, geopolitical, and historical variables, this study corroborates that crude oil trade link formation and dissolution are indeed political.

Among the six statistical learning models employed in this study, the Random Forest Classifier performs exceptionally well in all the train and test cycles, demonstrating that the model is quite robust in capturing the trends and technicalities of the dataset. This could be attributed to the ensemble approach of the model by creating multiple decision trees and reducing sampling bias. Valuable complementary insights have been revealed by the DL models, even though they were not able to surpass the benchmark performance of the Random Forest Model. The ANN model has been consistently outperformed by the CNN model in terms of stability, and this trend is particularly true for years of geopolitical volatility like 2008-09, 2014-15, and 2020-21, owing to the latter model's ability to capture feature dependencies and structural interactions across variables. Although ANN is able to grasp broader nonlinear trends, it shows greater volatility relative to CNN, thus corroborating the fact that CNN offers more stable, robust, and generalizable predictions.

Further, the models significantly gained from longer training windows, thus specifying the temporally embedded and path-dependent nature of crude oil trade linkages, along with the need for larger volumes of data for the models to perform well. Moreover, as suggested by observed patterns, a blend of economic fundamentals, geopolitical alignments, historical inertia, and policy interventions shapes crude oil trading dynamics, which constitutes a complex adaptive system rather than a rule-bound framework, underscoring the constraints of

traditional econometric models and emphasizing the increasing relevance of ML and DL models.

But it is quite essential to understand that no matter how robust a model is, given the intricacies of crude oil trading, it is still prone to unforeseen and unexpected global scenarios. The pronounced dives in the performance of the models around 2003, 2008, 2011, and 2014-15 corroborate this fact well; however, there is evidence that the models learn and adapt over time.

The capability of the models to capture non-linear patterns and temporal trade disruptions delivers actionable values for oil-exporting countries. To enable pre-emptive diplomatic engagement and logistical restructuring with an aim to maintain market access, the predictive success of the models can be utilized in recognizing trade partnerships at risk. Exporting nations can employ insights from turnover and substitution index trends to diversify contractual portfolios and reduce geopolitical overreliance on a restricted set of importers. The comprehensive analysis in this paper offers a framework for optimizing the timing and selection of trade re-engagement efforts in post-conflict environments, given that high reactivation periods in countries like Libya, Iraq, and Iran correspond to geopolitical normalization.

These predictive insights aid in forming strategies that ensure energy security and sustainability, along with diversification in terms of reinforcing relationships with existing and potential trade partners. Moreover, it stresses the importance of maintaining flexibility in framing geopolitical strategies that are linked to crude oil trade. Decision-makers can make use of the predictive power of the algorithms to predict potential conflicts and thus ensure robust and secure energy supply chains. Also, through the prediction of the formation and dissolution of crude oil trade links, the regulatory regimes can be made more flexible to incorporate the evolving arena of global trade relationships, which might entail revising treaties and agreements, sanctions, and qualitative bilateral and multilateral relations among countries.

Lastly, the increasing practical relevance of adaptive, data-driven prediction algorithms has been reinforced by these findings. Further research work can integrate ML and DL models by employing a hybrid framework for its capability to detect shifts in trade behavior and to assess supply chain vulnerabilities. Including a wide range of parameters affecting crude oil trading, which may encompass other diverse economic metrics like energy intensity of GDP, and import dependence ratio, political stability metrics, technological upgradations, environmental policies, and emerging trends of utilizing renewable energy resources in order to capture intricacies in an exhaustive way is also beneficial. This is all the more crucial in the face of an evolving global crude oil trading scenario under decarbonization efforts, geopolitical uncertainty, shifting alliances, and for anticipating disruptions in order to maintain a robust and forward-looking policy.

References

1) Bertoli, S., Goujon, M., & Santoni, O. (2016). The CERDI-seadistance database. *Études et Documents*, (7). CERDI. http://cerdi.org/production/show/id/1791/type_production_id/1
2) BP. (2021). *Statistical review of world energy 2021 – 70th edition*.
3) Cappelli, F., Carnazza, G., & Vellucci, P. (2023). Crude oil, international trade and political stability: Do network relations matter? *Energy Policy, 176*, 113479. https://doi.org/10.1016/j.enpol.2023.113479
4) Cen, Z., & Wang, J. (2019). Crude oil price prediction model with long short-term memory deep learning based on prior knowledge data transfer. *Energy, 169*, 160–171.
5) Colgan, J. D. (2013). Fueling the fire: Pathways from oil to war. *International Security, 38*(2), 147–180.
6) Deng, C., Ma, L., & Zeng, T. (2021). Crude oil price predict based on deep transfer learning: Shanghai crude oil as an example. *Sustainability, 13*(24), 13770.
7) Energy Institute. (2023). Statistical review of world energy: 72nd edition. Energy Institute.
8) Fischhendler, I., Herman, L., & Maoz, N. (2017). The political economy of energy sanctions: Insights from a global outlook 1938–2017. *Energy Research & Social Science, 34*, 62–71.
9) Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics, 81*, 709–727. https://doi.org/10.1016/j.eneco.2019.05.006
10) Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. http://www.deeplearningbook.org
11) Guan, Q., An, H., Gao, X., Huang, S., & Li, H. (2016). Estimating potential trade links in the international crude oil trade: A link prediction approach. *Energy, 102*, 406–415.
12) Hamilton, J. D. (2009). Causes and consequences of the oil shock of 2007–08. *Brookings Papers on Economic Activity*, *2009*(Spring), 215–261. Johns Hopkins University Press.
13) Han, X., Wang, L., Han, S. N., Chen, C., Crespi, N., & Farahbakhsh, R. (2015). Link prediction for new users in social networks. In *2015 IEEE International Conference on Communications (ICC)* (pp. 1250–1255). IEEE.
14) Jiao, X., Song, Y., Kong, Y., & Tang, X. (2021). Volatility predicting for crude oil based on text information and deep learning PSO-LSTM model. *Journal of Predicting, 41*(5), 933–944.
15) Kanchana, K., McLellan, B. C., & Unesaki, H. (2016). Energy dependence with an Asian twist? Examining international energy relations in Southeast Asia. *Energy Research & Social Science, 21*, 123–140.
16) K. Leetaru (2013). "The GDELT Project." https://www.gdeltproject.org/.
17) Li, X., Shang, W., & Wang, S. (2019). Text-based crude oil price predicting: A deep learning approach. *International Journal of Predicting, 35*(4), 1548–1560.
18) Liang, X., Luo, P., Li, X., Wang, X., & Shu, L. (2023). Crude oil price prediction using deep reinforcement learning. *Resources Policy, 81*, 103363. https://doi.org/10.1016/j.resourpol.2023.103363

19) Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology, 58*(7), 1019–1031.
20) Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87–106.
21) Nosratabadi, S., Mosavi, A., Duan, P., Ghamisi, P., Ferdinand, F., Band, S. S., Reuter, U., Gama, J., & Gandomi, A. H. (2020). Data science in economics: Comprehensive review of advanced machine learning and deep learning methods. *Mathematics, 8*(10), 1799.
22) Organization of the Petroleum Exporting Countries (OPEC). World proven crude oil reserves by country. OPEC Annual Statistical Bulletin, 2023. https://asb.opec.org/data/ASB_Data.php.
23) Ozturk, O. (2024). The impact of AI on international trade: Opportunities and challenges. *Economies, 12*(11), 298.
24) Patidar, A. K., Jain, P., Dhasmana, P., & Choudhury, T. (2024). Impact of global events on crude oil economy: A comprehensive review of the geopolitics of energy and economic polarization. *GeoJournal, 89*, 50.
25) Shaffer, B. (2011). *Energy politics*. University of Pennsylvania Press.
26) Shen, L., Bao, Y., Hasan, N., Huang, Y., Zhou, X., & Deng, C. (2024). Intelligent crude oil price probability predicting: Deep learning models and industry applications. *Computers in Industry, 163*, 104150.
27) Silva, T. C., Wilhelm, P. V. B., & Amancio, D. R. (2024). Machine learning and economic predicting: The role of international trade networks. *Physica A: Statistical Mechanics and its Applications, 649*, 129977. https://doi.org/10.1016/j.physa.2024.129977
28) Skeet, I. (1989). *OPEC: Twenty-five years of prices and politics*. Cambridge University Press.
29) Su, C.-W., Qin, M., Tao, R., & Moldovan, N.-C. (2019). Is oil political? From the perspective of geopolitical risk. *Defence and Peace Economics*.
30) Syropoulos, C., Felbermayr, G., Kirilakha, A., Yalcin, E., & Yotov, Y. V. (2023). The global sanctions data base–Release 3: COVID-19, Russia, and multilateral sanctions. *Review of International Economics, 32*, 12-48. https://doi.org/10.1111/roie.12691
31) Tuninetti, M., Tamea, S., Laio, F., & Ridolfi, L. (2017). To trade or not to trade: Link prediction in the virtual water network. *Advances in Water Resources, 110*, 528–537.
32) Voukelatou, V., Pappalardo, L., Miliou, I., Gabrielli, L., & Giannotti, F. (2020). Estimating countries' peace index through the lens of the world news as monitored by GDELT. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 216–225). IEEE.
33) Wakeford, J., & Swilling, M. (2014). Implications of increasing world oil scarcity for national food security in South Africa. *Agrekon: Agricultural Economics Research, Policy and Practice in Southern Africa, 53*(4), 68–91.
34) Wohlfarth, T., & Ichise, R. (2008). Semantic and event-based approach for link prediction. In T. Yamaguchi (Ed.), *Practical aspects of knowledge management: 7th International Conference, PAKM 2008, Yokohama, Japan, November 22–23, 2008, Proceedings* (Vol. 5345, pp. 50–63). Springer Berlin Heidelberg.

35) Wu, J., Dong, J., Wang, Z., Hu, Y., & Dou, W. (2023). A novel hybrid model based on deep learning and error correction for crude oil futures prices predict. *Resources Policy, 83*, 103602.
36) Zhang, C., Zhang, H., Yuan, D., & Zhang, M. (2016). Deep learning based link prediction with social pattern and external attribute knowledge in bibliographic networks. In *2016 IEEE International Conference on Internet of Things (iThings), IEEE Green Computing and Communications (GreenCom), IEEE Cyber, Physical and Social Computing (CPSCom), and IEEE Smart Data (SmartData)* (pp. 815–821).
37) Zheng, Y., Xu, Z., & Xiao, A. (2023). Deep learning in economics: A systematic and critical review. *Artificial Intelligence Review, 56*, 9497–9539.

Statements & Declarations:

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used Grammarly in order to tackle punctuation and grammatical errors. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Competing Interests

The author has no relevant financial or non-financial interests to disclose.

Funding

The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability Statement

The data and the codes used to generate the results are available from the corresponding author upon request.

Author Contributions

Ritu Singh conceived the study, collected and processed the data, designed and implemented the methodology, conducted the analysis, and wrote the manuscript.

Appendix A. Further details

Appendix A: Further details

A.1. Logistic Ridge Regression

One of the major issues faced by prediction-based models is overfitting. Overfitting is a phenomenon in which the statistical learning model instead of capturing the fundamental pattern of data, captures the inherent noise in it. The result is that though it fits the training set data very well, it fails to generalize and perform well on new and unseen data or the test data. This situation is in a way synonymous with rot learning where the student doesn't understand the content and just memorizes it. Precisely, this inconsistency arises because of the fact that the model is excessively complex and captures every fluctuation in the training dataset even the random fluctuations and outliers. This is related to the bias-variance trade-off in machine learning jargon. The more complex the model is, the lesser the bias, the greater the fit on the training data, and the greater the variance. Mathematically, the ridge regularization relates to minimizing the following loss function:

$$L(\beta)_{ridge} = -\frac{1}{N}\sum_{i=1}^{N}\left[T_{ijt}\, log\left(\hat{T}_{ijt}\right) + \left(1 - T_{ijt}\right) log\left(1 - \hat{T}_{ijt}\right)\right] + \lambda \sum_{k=1}^{p} \beta_{} \quad \text{Eq. (A.1)}$$

Where, $L(\beta)_{ridge}$ is the loss function associated with ridge regularization, λ is the regularization strength, p is the total number of predictors in the model, N is the number of observations and β is the coefficient of the model. The loss function or the cross-entropy loss function (or log loss) measures the discrepancy between the observed categories of the data and the predicted probabilities of these categories by the model. The first part $T_{ijt}.\log\left(\hat{T}_{ijt}\right)$ captures the loss when the actual outcome is positive i.e., $T_{ijt} = 1$ and the second part $\left(1 - T_{ijt}\right).\log\left(1 - \hat{T}_{ijt}\right)$ captures the loss when the actual outcome is negative i.e., $T_{ijt} = 0$.

Each term penalizes the divergence of the predicted probability from the actual value. When the anticipated probability closely aligns with the predicted result, the penalty is minimal. In cases where there is a considerable deviation between the anticipated likelihood and the observed outcome (for instance, projecting a high likelihood when the observed outcome is zero, or vice versa), the penalty is significant. We augment this loss function with an L2 regularization term $\lambda \sum_{k=1}^{p} \beta_k^2$. It is added to prevent overfitting by penalizing large coefficients. λ controls the strength of this penalty; the higher the λ, the more the model is penalized for having large coefficients, encouraging the model to be simpler and, ideally, more generalizable. The goal of the logistic ridge regression model is to minimize $L(\beta)_{ridge}$ by finding appropriate model coefficients.

Every machine learning model's performance is crucially dependent on its hyperparameters that are tuned using the dataset by employing a cross-validation approach. In the case of Logistic Ridge Regression, this hyperparameter is λ which is the regularization strength or the penalty term. In this research, λ is chosen by making use of grid search combined with time series cross-validation. Grid search is an extensive exploration method that assesses the performance of machine learning models based on a predefined grid of hyperparameters. This study has established a grid of probable λ values capturing varying degrees of penalties. The method undertakes each λ value from the established grid, trains it on the training dataset, and evaluates the performance using the test dataset based on a particular evaluation metric -

accuracy (the ratio of number of correct predictions to total number of predictions) in this study. The main aim is to choose the λ value that results in the most accurate performance and thus provides a balance between model complexity and generalizability.

Contrary to the traditional 5-fold cross-validation technique, time series cross-validation is a sophisticated approach that keeps the temporal nature of the dataset in mind. Validation in this case is exclusively carried out on the data immediately succeeding the training dataset in time. This process is sequential and progresses in time. It goes on enlarging the training dataset chronologically and shifts the validation set or test set accordingly. The model's performance with each λ value is accessed using the accuracy metric in each temporal fold (this study uses a 5-fold time series cross-validation split) ensuring adaptability and consistency over time. By integrating grid search with time series cross-validation, this study adopts a methodologically sound and rigorous approach for accommodating the complexity of the model and nuances of the chronological nature of the dataset under consideration.

A.2. Random Forest Classifier

A decision tree is a model in the machine learning paradigm used for predicting concerned variables. It is a graphical representation comprising nodes and branches. Each node signifies a choice between different features of the dataset, and each terminal node gives the final result of classification i.e., whether there will be a crude oil trade link or not based on the features of the dataset. Fig. A.1 shows a simplified representation of a decision tree incorporating only two variables from the dataset as features – number of treaties and Avgtone between the trading country pairs at a given time. Focussing on the root of the decision tree which is the starting point, V2_tone has been used indicating that the tone of events occurring between country pairs is the first and foremost feature this model considers as prominent in predicting the formation or cessation of trade links.

If the V2_tone value is less than or equal to 3.37, this model proceeds along the left branch of the decision tree, otherwise, it will move toward the right branch. For instance, let us move towards the left branch: if the number of treaties between trading country pairs is less than or equal to 235.5 and the V2_tone is less than or equal to -1.51, this decision tree reaches the leftmost terminal node named node #3 and predicts that there will be no crude oil trade link between the country pair under consideration. This prediction is based on the fact that owing to the recursive binary splitting, the class distribution of node #3 is such that 60% of the instances belong to no trade category and 40% of the instances belong to the trade category hence based on the majority category, this node predicts that there will be no trade link. Further, node #3 contains 18.1% of the total training samples. Each branch carries on this process of classification based on the specified cut-off thresholds until a leaf or terminal node gives the result. This series of choices is based on what the model has learned from the particular dataset at hand during the model training process.

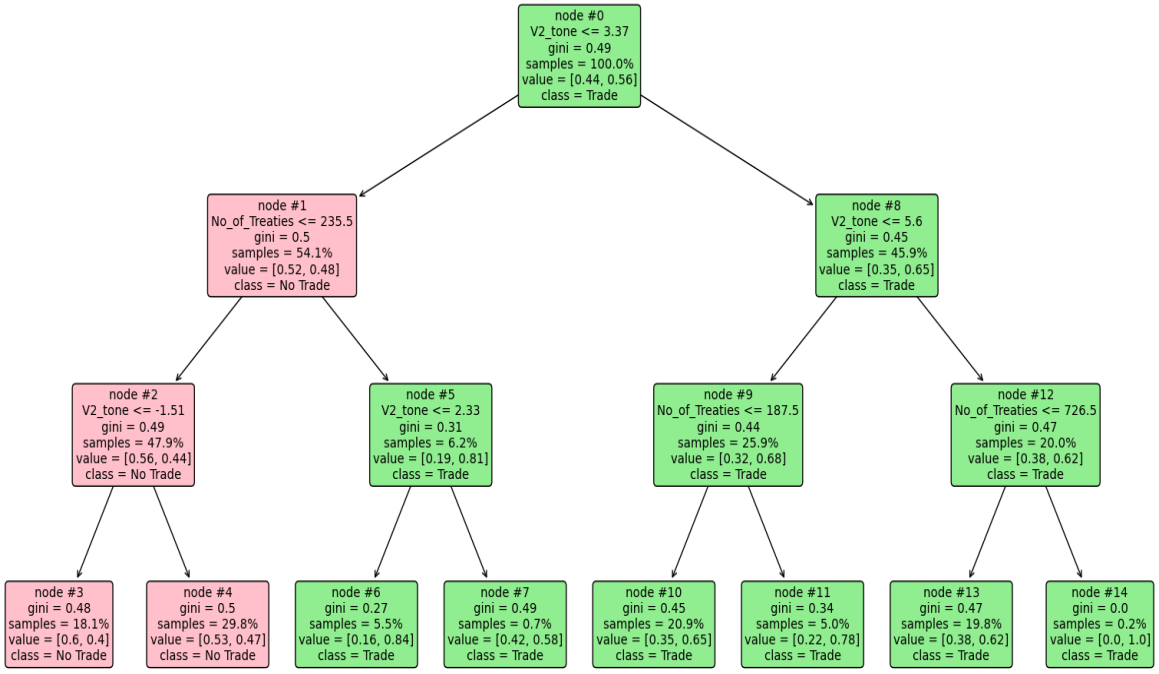


Fig. A.1. A simple decision tree

One question that naturally arises is, what are the criteria for choosing between the features to be used for splitting at a particular node? This research uses the Gini index as the main criterion. It is a measure of the purity of the dataset and is usually employed in the case of classification problems to select features that most effectively split the data. For a binary classification task, the value of the Gini index hovers in the range of 0 to 0.5 where a value of 0 indicates that there is no impurity in the dataset and all the instances uniformly belong to only one of the two categories whereas a value of 0.5 indicates that 50% of the instances belong to category 1 and 50% belong to category 0. Suppose there is a dataset D having K classes or categories, then the Gini impurity (G) of D is calculated using the following equation:

$$G(D) = 1 - \sum_{i=1}^{K} p_i^2 \qquad \text{Eq. (A.2)}$$

In the above formulation, p_i represents the proportion of observations that belong to class i in the dataset D. Furthermore, when D is split into two subsets D_1 and D_2 in a decision tree based on a feature F and a threshold T, the reduction in Gini impurity pertaining to this division of the dataset is given by:

$$\Delta G(D, F, T) = G(D) - \left(\frac{D_1}{D} G(D_1) + \frac{D_2}{D} G(D_2)\right) \qquad \text{Eq. (A.3)}$$

The partition that leads to the maximal decrease in the Gini index ΔG based on comparing ΔG corresponding to all the features of the dataset is selected as the optimal and final way to divide the given dataset at each node of the decision tree. This process is replicated in a recursive way for each subsequent node until a stopping criterion is met, such as reaching a

maximum tree depth, minimum number of samples, or emphasizing a predefined threshold for the minimum decrease required in the Gini index for proceeding with a split.

The Random Forest Classifier ameliorates the decision tree algorithm through the creation of an ensemble of decision trees commonly trained with the 'bagging' method which is used to reduce variance within a noisy dataset. The process begins with the creation of multiple bootstrapped (sampling with replacement) data from the original dataset under consideration. Each sampled dataset is then used to train a unique decision tree ensuring that each tree learns from a different subset of data and therefore reduces the sampling bias and encourages diversification in the learning process of the model. Further randomness in the process is introduced by choosing a random subset of features at each split in the decision tree in contrast to utilizing the entire set of features which guarantees that decision trees are not correlated and hence reduces the variance of the model owing to the fact that no single feature dominate the entire decision-making process. For the final classification task, the prediction of cessation or formation of crude oil trade link is made by majority voting in which each decision tree votes for one of the two classes, and the category that receives a majority of votes is selected as the final prediction of the Random Forest Classifier model.

Even after all the randomness introduced in the process, this classifier is still prone to model overfitting but not as much as other models. To counter the risk of overfitting, we usually specify some kind of stopping criterion that determines when a tree stops growing. This is usually achieved by tuning the hyperparameters of the random forest classifier that control the learning process and model complexity.

A.3. Support Vector Machine (SVM)

The optimization function with constraints for a SVM model is given as:

$$\min_{\boldsymbol{W},b} \frac{1}{2}\boldsymbol{W}^T\boldsymbol{W} + C\Sigma\xi_{ijt}$$

$$\text{s,t. } T_{ijt}\left(\boldsymbol{W}^T\phi\left(\boldsymbol{X}_{ijt}\right) + b\right) \geq 1 - \xi_{ijt}, \quad \forall i,j,t \qquad \text{Eq. (A.4)}$$

$$\xi_{ijt} \geq 0$$

The minimization of $\boldsymbol{W}^T\boldsymbol{W}$ instead of $\boldsymbol{W}$ assures that the optimization problem is convex so that the model can achieve a global minimum and is not stuck in some local minima. By minimizing $\frac{1}{2}\boldsymbol{W}^T\boldsymbol{W} + C\Sigma\xi_{ijt}$, the SVM algorithm is attempting to find the smallest possible weight vector $\boldsymbol{W}$ that correctly predicts the points in the training dataset and the minimization of $\boldsymbol{W}$ corresponds to the maximization of the margin between the hyperplane and the nearest classified data point because the margin is inversely proportional to the weight vector of the hyperplane.

The expression $\boldsymbol{W}^T\phi\left(\boldsymbol{X}_{ijt}\right)$ is the dot product of the weight vector of the hyperplane with the transformed feature vector stipulating the orientation of a particular data point relative to the hyperplane in the transformed feature space. The constraint $T_{ijt}\left(\boldsymbol{W}^T\phi\left(\boldsymbol{X}_{ijt}\right) + b\right) \geq 1 - \xi_{ijt}$ specifies that each data point is classified correctly by being on the correct side of the hyperplane given by its class label T_{ijt} depicting the existence or non-existence of crude oil trade links beyond the margin given by $1 - \xi_{ijt}$ or at least within an allowable distance from

the margin determined by the slack variable ξ_{ijt}. The term X_{ijt} is a vector of features pertaining to the crude oil trading country pair i (exporting country) and j (importing country) at time t. In the case of the SVM model, the support vectors defined by the points lying on the edge of the margin have a critical role in optimization as these are the most relevant points that the hyperplane must classify correctly while model training and hence, the orientation and position of the hyperplane in the feature space essentially depends on these data points.

The final prediction for the class label based on the feature vector of a particular data point is given by the following decision function:

$$f(\boldsymbol{X}_{ijt}) = (\boldsymbol{W}^T \phi(\boldsymbol{X}_{ijt}) + b \qquad \text{Eq. (A.5)}$$

If, $f(\boldsymbol{X}_{ijt}) > 0$, it means the crude oil trade link will exist, and if $f(\boldsymbol{X}_{ijt}) < 0$, it implies that the crude oil trade link won't exist between country pairs i and j at time t.

Again, the entire architecture of the SVM model like any other machine learning model revolves around its hyperparameters that need to be carefully deliberated using cross-validation techniques for training the model robustly. These hyperparameters play an extensive role in controlling overfitting and enhancing the performance of the model on the given dataset.

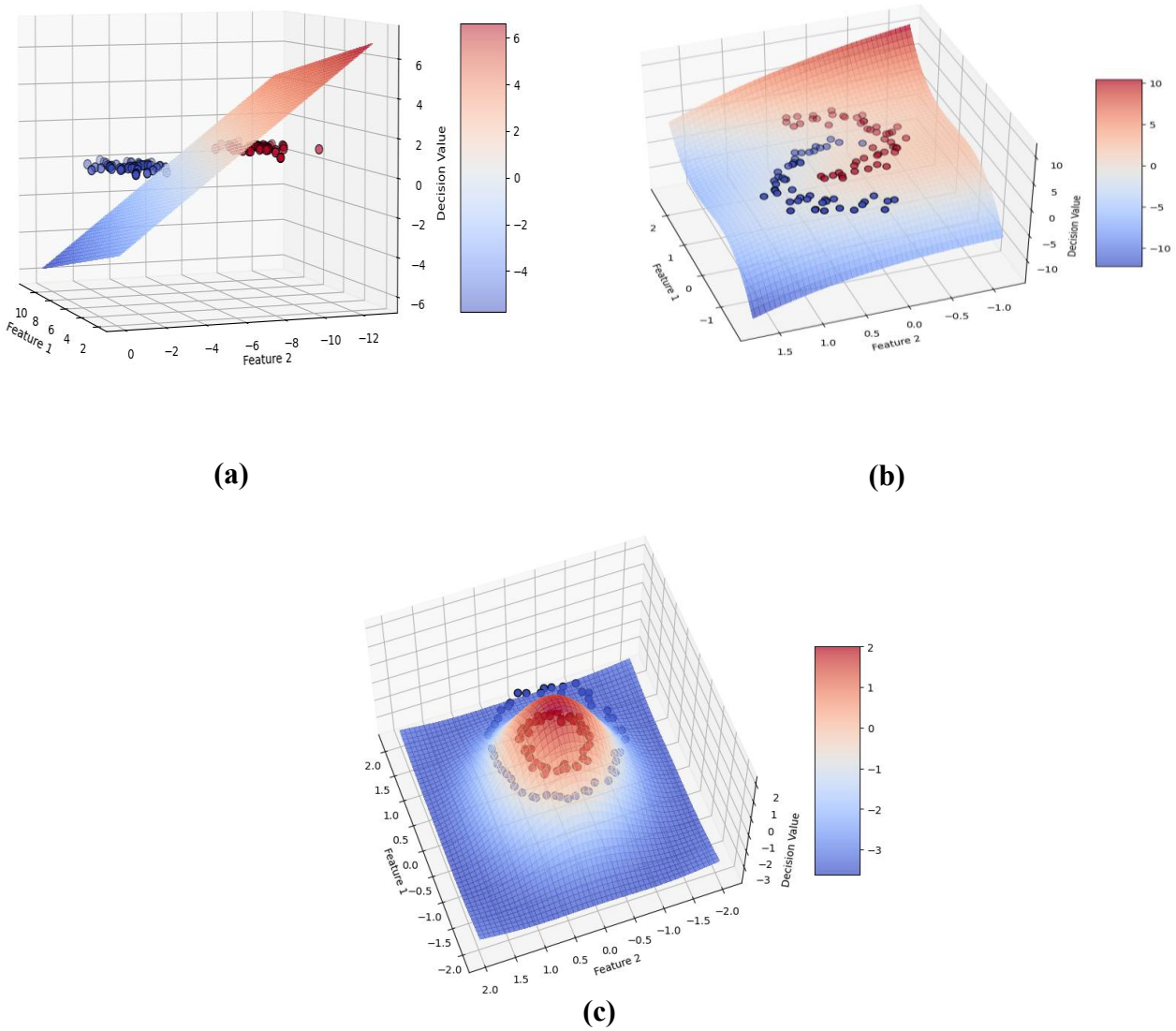
First and foremost is the 'Kernel function' which is the function used for the transformation of features to a higher-dimensional space for aiding linear separation. Technically, it renders the model another viewpoint to observe data points from various angles and thus helps in revealing underlying intricate patterns. The choice of kernel type – linear, polynomial, or Radial Basis Function (RBF) mandates the degree of flexibility in the model for defining the decision boundary, the more flexible the model, the lesser the bias, the larger the variance, and the greater is the risk of overfitting owing to the bias-variance tradeoff.

The regularization parameter C, denotes the strength of penalty for overfitting of the model and administers the balance between model accuracy and generalizability. A C value that is very small leads to a model with a wider margin and therefore, permits more misclassifications of the data points within the dataset and thus bolsters generalizability. On the other hand, a larger C value produces a narrower margin which renders more accurate classification of training set data points but is much more prone to overfitting.

Another prominent hyperparameter is the 'gamma' value which comes into effect when employing an SVM with RBF kernel and determines the influence radius of a single training data point. A lower 'gamma' value pertains to smoother decision boundaries and the model fits less closely to the training data set ensuring the generalizability of the model whereas, a higher 'gamma' value fits tightly to the training data set thus ensuring less bias but the model has a higher risk of overfitting. Furthermore, in the case of polynomial kernels, the degree of polynomial is a hyperparameter controlling the model complexity of the decision boundary employed. In this research, these hyperparameters have been chosen by employing grid search combined with time series cross-validation.

Fig. A.2 provides a rudimentary visualization of how SVM works based on hypothetical datasets in which two features are plotted along the X and Y axes, while the Z axis denotes the SVM classification model's decision boundary or the hyperplane which is also a measure of the certainty of the model's prediction, specifically, the farther a data point is to the

hyperplane, the more confidence we have regarding the classification of that data point while training the model. A decision value of 0 marks the decision boundary, portraying the model's calculation of the optimal hyperplane dividing the dataset into two categories. Points on one side of this decision boundary are predicted as belonging to class 1, whereas points on the opposite side are classified as belonging to class -1. The SVM model is trained on the concerned dataset in order to find the decision boundary or hyperplane that maximizes the margin or the distance between the boundary and the nearest data points from each class, known as the support vectors.



(a) **(b)**

(c)

Fig. A.2. A simple visualization of SVM with linear and non-linear decision boundaries: part (a) shows a linear kernel, part (b) shows a polynomial kernel and part (c) shows a radial kernel technically termed RBF (Radial Basis Function)

A.4. Artificial Neural Network (ANN)

The ANN model is comprised of three main layers – input layer, hidden layer, and output layer as depicted by Fig. A.3. Given the dataset under consideration, let the input features for i^{th} data point be represented as a vector $\boldsymbol{x}_i$. Assuming that the dataset has N observations

(trade links), all the dataset of input features can be represented as a matrix **X** of size N×m where m is the number of features.

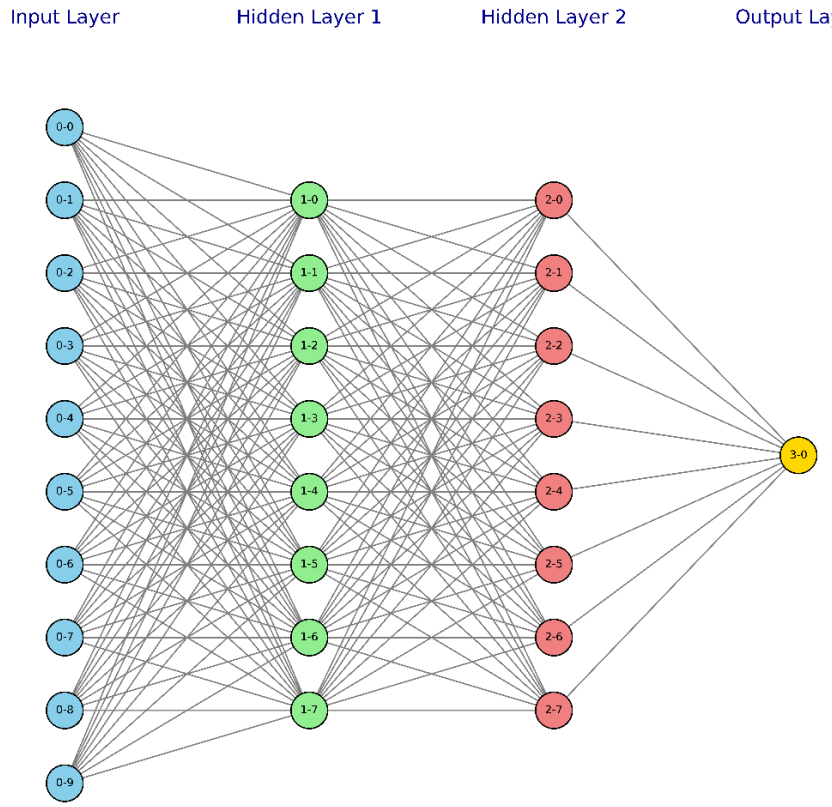


Fig. A.3. A simple visualization of the ANN model

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ . \\ . \\ . \\ \boldsymbol{x}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{11} & \boldsymbol{x}_{12} & . & . & . & \boldsymbol{x}_{1m} \\ \boldsymbol{x}_{21} & \boldsymbol{x}_{22} & . & . & . & \boldsymbol{x}_{2m} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \boldsymbol{x}_{N1} & \boldsymbol{x}_{N2} & . & . & . & \boldsymbol{x}_{Nm} \end{pmatrix} \qquad \text{Eq. (A.6)}$$

The target vector **Y** for all the data points can be written as:

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ . \\ . \\ . \\ \boldsymbol{y}_N \end{pmatrix} \qquad \text{Eq. (A.7)}$$

So, the input layer will consist of m neurons corresponding to the m input features. The model processes one observation or a batch of observations at a time depending on the implementation framework. Each of these observations having m features is passed through the network when one observation at a time is used, otherwise, if batch processing is used, the network still has m input neurons but it processes multiple observations simultaneously.

Assume that there are L hidden layers, where each layer l contains m_l neurons, and the output layer will represent a single neuron having the binary classification outcome.

The forward propagation involves applying a linear transformation to each hidden layer l followed by a non-linear activation function. This operation can be represented for the first hidden layer $l = 1$ as:

$$\boldsymbol{h}^{(1)} = f\left(\boldsymbol{X}\boldsymbol{W}^{(1)} + \boldsymbol{b}^{(1)}\right) \quad \text{Eq. (A.8)}$$

Where, $\boldsymbol{W}^{(1)}$ is the weight matrix of size m$\times m_1$ connecting the input layer to the first hidden layer, $\boldsymbol{b}^{(1)}$ is the bias vector of size $1\times m_1$, and $f(\cdot)$ is the activation function being applied element-wise. The output of each neuron j in the first hidden layer $l = 1$ prior to applying the activation function is basically a weighted sum of all the inputs from the input layer along with a bias term and can be written as:

$$z_j^{(1)} = \sum_{k=1}^{m} W_{kj}^{(1)} x_{ik} + b_j^{(1)} \quad \text{Eq. (A.9)}$$

Here, $z_j^{(1)}$ is the input to the activation function of the jth neuron in the first hidden layer, $W_{kj}^{(1)}$is the weight connecting the kth input feature x_{ik} to the jth neuron in the first hidden layer, $b_j^{(1)}$ is the bias term associated with jth neuron in the first hidden layer, and x_{ik} is the kth feature of the ith observation.

Further, the output from the first hidden layer $h^{(1)}$ becomes the input for the next hidden layer. In general, for a layer l, the output can be represented as:

$$\boldsymbol{h}^{(l)} = f\left(\boldsymbol{h}^{(l-1)}\boldsymbol{W}^{(l)} + \boldsymbol{b}^{(l)}\right) \quad \text{Eq. (A.10)}$$

Finally, the output layer applies the sigmoid function $\sigma(\cdot)$ in order to produce a probability for the binary classification (SoftMax function is used in case of multiple categories):

$$\hat{y} = \sigma\left(\boldsymbol{h}^{(L)}\boldsymbol{W}^{(L+1)} + b^{(L+1)}\right) \quad \text{Eq. (A.11)}$$

Where, $\hat{y}$ is the predicted output, $\boldsymbol{W}^{(L+1)}$ is the weight matrix connecting the last hidden layer to the output layer (size $m_L \times 1$), and $b^{(L+1)}$ is the bias to the output layer which is a scalar. Prior to employing the sigmoid function, the ANN model's output for the present task at hand is a single scalar value representing the raw prediction also termed the logit. This raw output is actually a linear combination of the outputs of the last hidden layer. The logit is then passed through the sigmoid function to produce a probability between 0 and 1.

The model tends to minimize the binary cross entropy loss function for which the process of backpropagation is employed. It consists of computing the gradients of the loss function with respect to the weights and biases, and updating them using gradient descent algorithms:

$$\boldsymbol{W}^{(l)} \Leftarrow \boldsymbol{W}^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^{(l)}} \quad \text{Eq. (A.12)}$$

$$\boldsymbol{b}^{(l)} \Leftarrow \boldsymbol{b}^{(l)} - \eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{b}^{(l)}} \quad \text{Eq. (A.13)}$$

Where, η is the learning rate. This weight updating process is repeated iteratively until the model converges to a set of weights that minimizes the loss function and hence makes predictions of the 'Trade_link' variable.

The performance of machine learning models and in particular deep learning models depends crucially on the hyperparameters that are set before the training process begins and are not learned from the data. With respect to the ANN model, the common hyperparameters include – number of layers (L) including both the hidden layers and the output layer, number of neurons per layer (m_l) representing number of neurons in each hidden layer, activation function referring to the non-linear functions applied to the output of each neuron such as ReLU, Sigmoid, Tanh, or others, learning rate (η) depicting the step size used by the optimization algorithms to update the model's weights during training process, batch size which involves number of training examples used in one iteration to update weights of the model, number of epochs capturing the number of times the entire training dataset is passed through the network during training process, dropout rate (fraction of neurons randomly dropped during training to prevent overfitting), regularization parameters (typically L2 regularization or weight decay), and optimizer (optimization algorithm used to minimize the loss function such as Adam, RMSprop, SGD or others).

Owing to the temporal dependencies in the dataset, the traditional cross-validation methods are not appropriate here. Therefore, this research has utilized the power of the time series cross-validation approach to tune the hyperparameters of the model. In this case, the dataset is split into training and validation sets in a way that the model is always trained on the past data and validated on the future data. For this, the rolling origin or expanding window approach has been followed where training starts on a small initial set of observations, and validation is performed on the next time step. The training set is then expanded by including the next time step, and validation is performed on the subsequent time step. This process has been combined with grid search to find the optimal hyperparameters with accuracy being the evaluation metric.

A.5. Convolutional Neural Network (CNN)

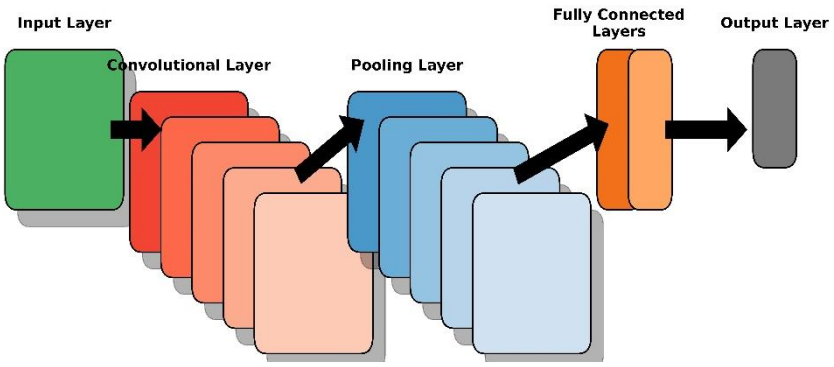


Fig. A.4. A simple visualization of the CNN model

Fig. A.4. demonstrates the typical structure of a CNN model. The feature data is fed into the network via the input layer. Each observation is presented as an input matrix with dimensions

$1 \times m$, where m is the number of features. The convolutional layer is tasked with applying a set of filters to the input data, which are used to detect spatial or local patterns such as interactions between different features over time. The filters move across the input data and thus extract feature maps that capture the essential patterns. The red blocks show the different filters applied to the input matrix, in which each red block represents a different filter.

The function of the pooling layer is to reduce the spatial dimensions of the feature maps generated by the convolutional layer, usually by taking the average or maximum value within a region. This layer down samples the data and thus reduces the computational load and makes the network robust to distortion in the input data. Each blue block in the diagram represents a pooled version of the feature maps from the previous convolutional layer, and the outcome after convolution and pooling goes through the fully connected layers in which each neuron is connected to every neuron in the previous layer. The output of this layer is fed into the sigmoid function (for binary classification) or SoftMax function (for multiple categories). The output layer finally provides the probability that a trade link exists between the exporting and importing countries.

Let $\boldsymbol{x}_i$ represent the input tensor for the ith observation and can be expressed as:

$$\boldsymbol{x}_i = \begin{bmatrix} x_i^{(1)} & x_i^{(2)} & . & . & . & x_i^{(m)} \end{bmatrix} \quad \text{Eq. (A.14)}$$

where, $x_i^{(m)}$ is the value of the mth feature for ith observation corresponding to a specific exporting country, importing country and year. The input matrix is reshaped to have three dimensions:

$$\boldsymbol{X} \rightarrow (N, m, 1) \quad \text{Eq. (A.15)}$$

Here, N is the number of observations, m is the number of features, and 1 represents the channel dimension in order to add extra dimension for the data to be compatible for employing CNN model (analogous to depth in case of image recognition). The convolutional layer employs a set of filters or kernels across the input data. Suppose that we have a filter size of 2:

$$\boldsymbol{W}_k = \begin{bmatrix} w_k^{(1)} & w_k^{(2)} \end{bmatrix} \quad \text{Eq. (A.16)}$$

where, $w_k^{(1)}$ and $w_k^{(2)}$ are the weights of the filter. At each position, the filter multiplies its weights by the corresponding input values and sums the results. This operation is done for each possible position of the filter on the input vector. For example, the filter is applied to the first two elements of the input vector and its output is:

$$z_{i,1}^{(k)} = w_k^{(1)} \cdot x_i^{(1)} + w_k^{(2)} \cdot x_i^{(2)} + b_k \quad \text{Eq. (A.17)}$$

where, $z_{i,1}^{(k)}$ is the first element of the output feature map applied to the ith observation for kth filter, b_k is the bias term associated with the kth filter. After the filter has slid over the entire input vector, it produces a feature map $\boldsymbol{Z}_i^{(k)}$. This filter is applied to all the N observations in the dataset, generating a separate feature map for every observation. The resultant feature maps corresponding to the entire dataset are:

$$\boldsymbol{Z}^{(\kappa)} = \begin{bmatrix} \boldsymbol{Z}_1^{(k)} & \boldsymbol{Z}_2^{(k)} & \boldsymbol{Z}_3^{(k)} & . & . & \boldsymbol{Z}_N^{(k)} \end{bmatrix} \quad \text{Eq. (A.18)}$$

Here, $\boldsymbol{Z}^{(\kappa)}$ is the set of feature maps generated by the kth filter for all the observations, and $\boldsymbol{Z}_1^{(k)}$ is the feature map for the ith observation from the kth filter. Therefore, the output for one filter k will have the shape (N, m'), in which m' corresponds to the length of each feature map (depends on filter size, stride, and padding). In the case of K filters, each of these filters will produce a separate feature map for all N observations, so the final result after applying all filters will be a 3D tensor of shape (N, m', K).

After the convolution operation, an activation function is employed to introduce non-linearity, and then pooling is applied to reduce the dimensionality of the feature maps in order to make the model more efficient and robust. The next step is the flattening of feature maps (converting 2D feature map into a 1D vector by concatenating all the values having dimensions as (N, K×m')) which are then directly fed into the fully connected layers in which each neuron in the layer is connected to every output from the previous layer:

$$z_i = \boldsymbol{W}_{fc} \cdot \boldsymbol{p}_i + b_{fc} \qquad \text{Eq. (A.19)}$$

Here, $\boldsymbol{p}_i$ is the flattened vector from the pooled feature maps for ith observation, $\boldsymbol{W}_{fc}$ (vector representing weights of different pooled features for ith observation), and b_{fc} are the weights and bias of the fully connected layer respectively. If the data is more complex then activation function can be applied in the dense layer too. The final output layer employs a sigmoid function to produce the probability of trade link existence.

$$\hat{y}_i = \sigma(z_i) = \frac{1}{1 + e^{-z_i}} \qquad \text{Eq. (A.20)}$$

The hyperparameters of the CNN model have been tuned using grid search combined with time-series cross-validation with an expanding window approach. The crucial hyperparameters of the CNN model include learning rate, batch size, number of epochs, number of filters/kernels, stride (the number of elements the filter moves across the input matrix), padding (helps in preserving the spatial dimensions, and to get around with information loss at the borders, pooling size and type (commonly max pooling or average pooling), dropout (some random neurons are dropped during each forward pass of training i.e., their weights are temporarily set to 0) and L2 regularization (also called weight decay, adds a penalty to the loss function, activation function, and optimizer.

A.6. Evaluation metrics

A.6.1. Accuracy

The accuracy of a statistical learning model is expressed as the ratio of the number of correctly predicted data points to the total number of data points in the dataset. Precisely,

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{Eq. (A.21)}$$

Here, TP, TN, FP, and FN imply true positives, true negatives, false positives, and false negatives respectively. It captures the overall correctness of the model. But this can be misleading when the dataset is not balanced meaning that the distribution of both the categories is not equal in proportion so if a model predicts the majority class, it will still have a high accuracy given the fact that a major portion of the dataset belongs to that class.

Moreover, this metric doesn't differentiate between the types of errors: false positives and false negatives.

A.6.2. F1 Score

The F1 score of a machine learning model is given by the harmonic mean of precision and recall metrics and is prominent for cases where the classes are imbalanced. It is given by:

$$\text{F1 score} = 2*(\frac{Precision*Recall}{Precision+Recall}) \quad \text{Eq. (A.22)}$$

Here, precision and recall of the model employed are respectively defined as:

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{Eq. (A.23)}$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{Eq. (A.24)}$$

The term precision captures the accuracy of positive predictions (in this research, the existence of trade links) made by the classification model, and assesses the fraction of positive predictions made by the trained model that were correct. On the other hand, recall encapsulates the competence of the model to correctly identify actual positives thus giving the fraction of actual positives that were identified by the model. There is often a trade-off between precision and recall because, in essence, precision focuses on minimizing type I error (false positives) whereas, recall focuses on minimizing type II error (false negatives).

A.6.3. Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

To measure the ability of a machine learning model in terms of discriminating between the two categories, AUC is the ideal metric. A ROC (Receiver Operating Characteristic) curve portrays the ability of a binary classifier to discriminate between the two classes as the discrimination threshold is varied and is constructed by plotting the True Positive Rate (TPR or recall) corresponding to the False Positive Rate (FPR) at various threshold settings. Technically,

$$\text{TPR} = \frac{TP}{TP+FN} \quad \text{Eq. (A.25)}$$

$$\text{FPR} = \frac{FP}{TN+FP} \quad \text{Eq. (A.26)}$$

The AUC or area under the ROC curve exemplifies the likelihood that a machine learning classifier model will rank a randomly chosen positive instance higher than a randomly chosen negative instance. A model that is able to discriminate between the two categories perfectly has an AUC of 1, whereas a model with no discriminative ability has an AUC of 0.5 which is as good as random guessing and no predictive power.

A.7. Figures and tables

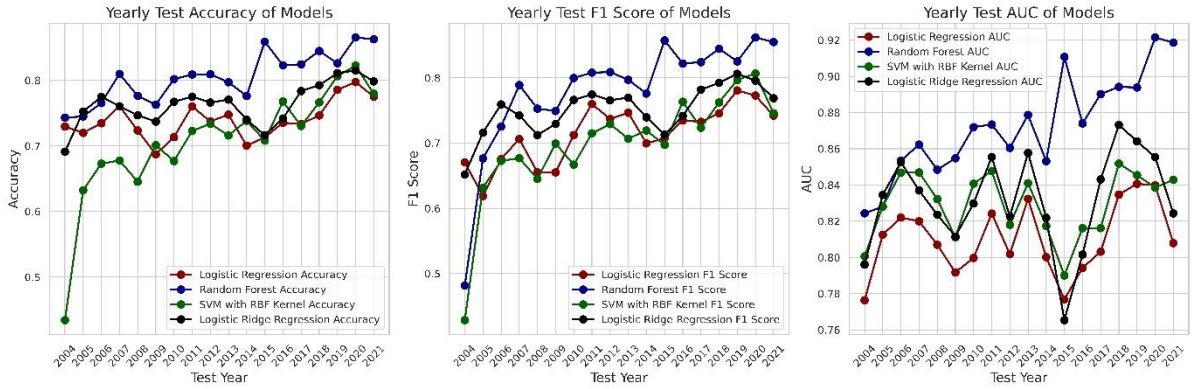


Fig. A.5. Accuracy, F1 score, and AUC plots for Logistic Regression, Logistic Ridge Regression, Random Forest classifier, and SVM for two-year train and one-year test cycles

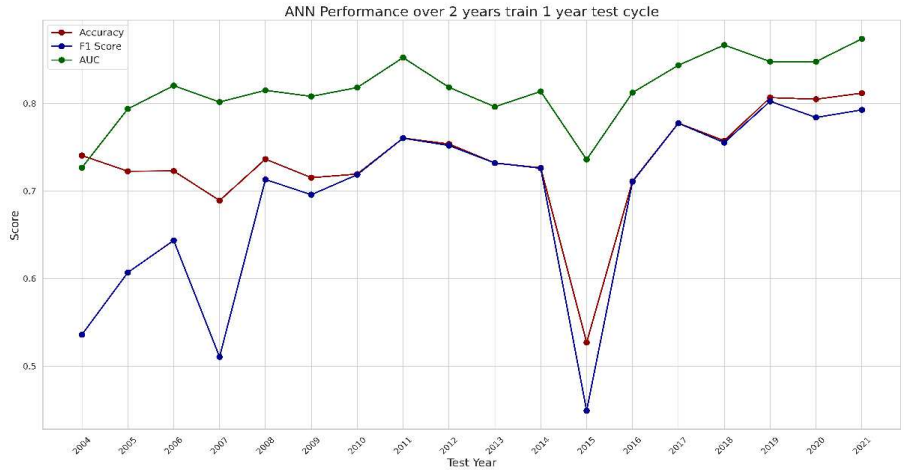


Fig. A.6. Accuracy, F1 score, and AUC plots for ANN for two-year train and one-year test cycles



Fig. A.7. Accuracy, F1 score, and AUC plots for CNN for two-year train and one-year test cycles

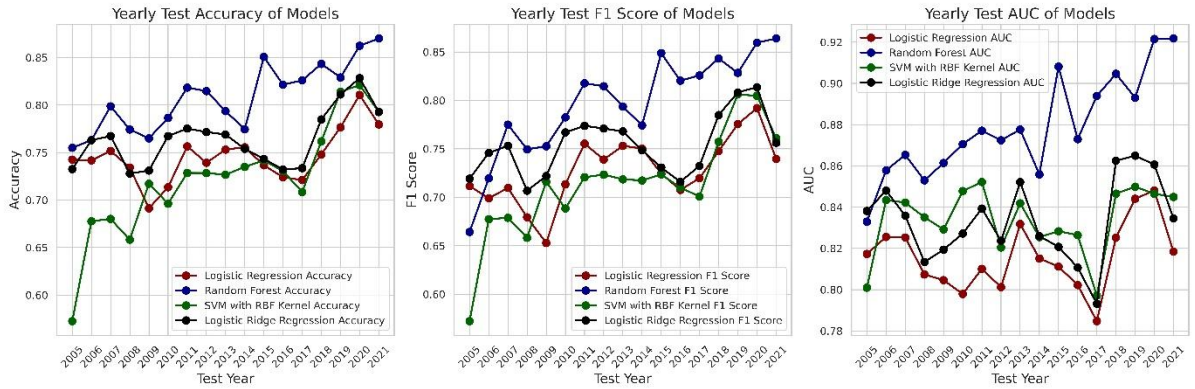


Fig. A.8. Accuracy, F1 score, and AUC plots for Logistic Regression, Logistic Ridge Regression, Random Forest classifier, and SVM for three-year train and one-year test cycles

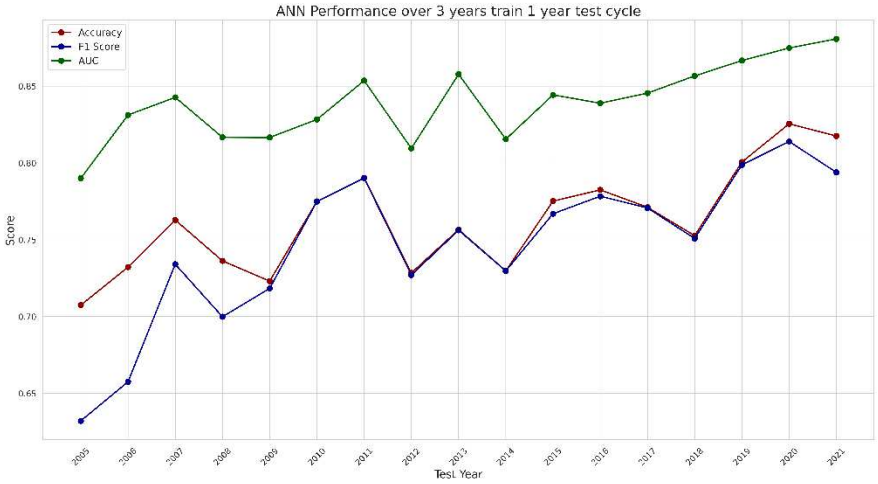


Fig. A.9. Accuracy, F1 score, and AUC plots for ANN for three-year train and one-year test cycles

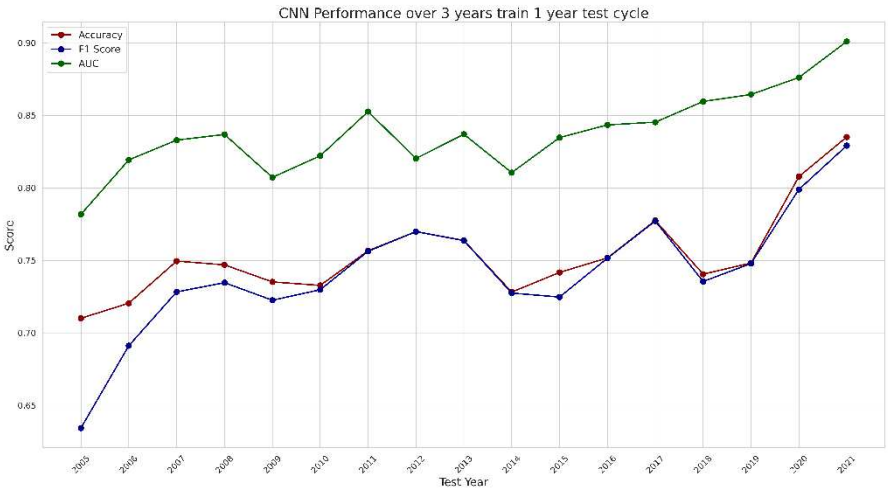


Fig. A.10. Accuracy, F1 score, and AUC plots for CNN for three-year train and one-year test cycles

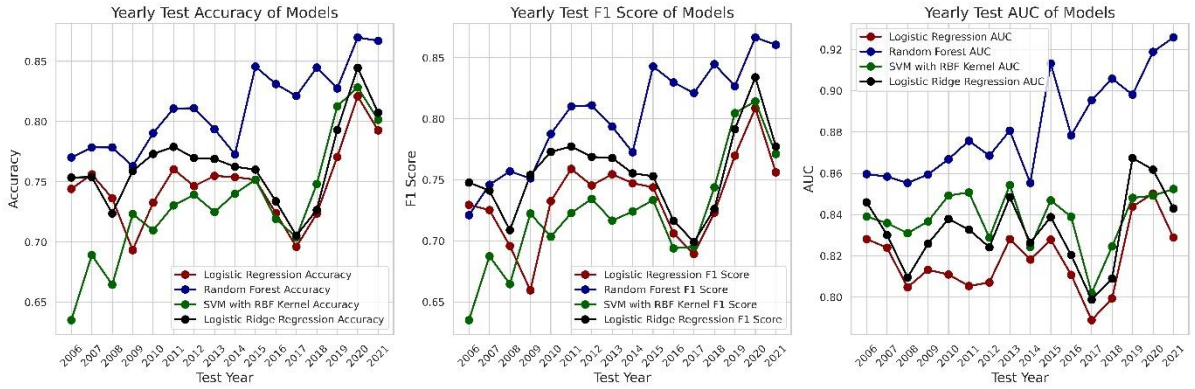


Fig. A.11. Accuracy, F1 score, and AUC plots for Logistic Regression, Logistic Ridge Regression, Random Forest classifier, and SVM for four-year train and one-year test cycles

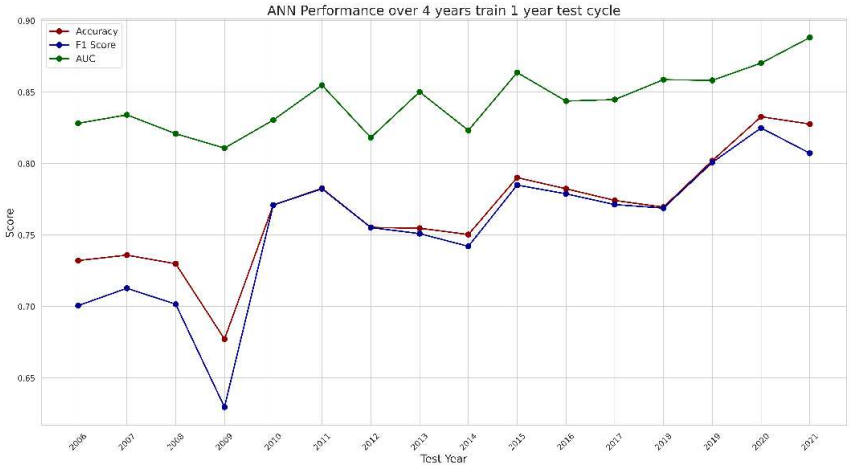


Fig. A.12. Accuracy, F1 score, and AUC plots for ANN for four-year train and one-year test cycles

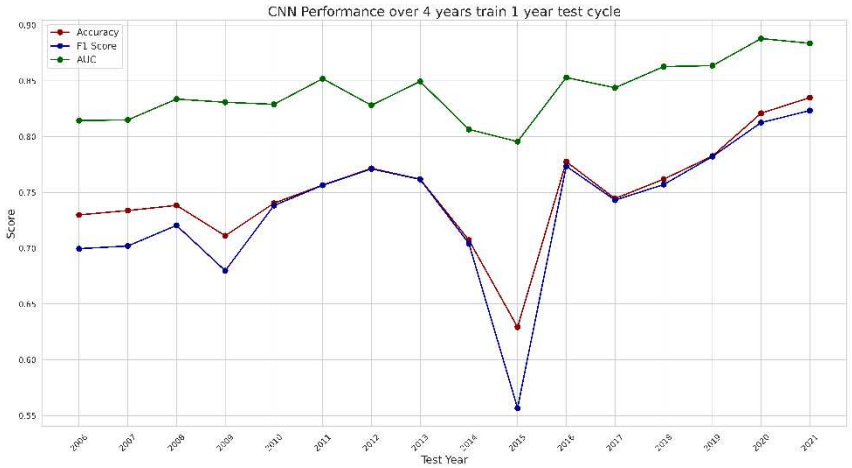


Fig. A.13. Accuracy, F1 score, and AUC plots for CNN for four-year train and one-year test cycles

Table A.1. Average test scores for one-year train and one-year test cycles

Model	Accuracy	F1 Score	AUC
Logistic Regression	0.73	0.67	0.79
Logistic Ridge Regression	0.76	0.72	0.82
Random Forest	0.80	0.76	0.86
SVM with RBF Kernel	0.70	0.69	0.81
ANN	0.74	0.69	0.80
CNN	0.75	0.71	0.81

Table A.2. Average test scores for two-year train and one-year test cycles

Model	Accuracy	F1 Score	AUC
Logistic Regression	0.74	0.71	0.81
Logistic Ridge Regression	0.76	0.75	0.83
Random Forest	0.81	0.78	0.87
SVM with RBF Kernel	0.71	0.70	0.83
ANN	0.73	0.69	0.82
CNN	0.73	0.71	0.83

Table A.3. Average test scores for three-year train and one-year test cycles

Model	Accuracy	F1 Score	AUC
Logistic Regression	0.74	0.73	0.82
Logistic Ridge Regression	0.76	0.75	0.83
Random Forest	0.81	0.80	0.88
SVM with RBF Kernel	0.72	0.71	0.83
ANN	0.76	0.75	0.84
CNN	0.75	0.74	0.84

Table A.4. Average test scores for four-year train and one-year test cycles

Model	Accuracy	F1 Score	AUC
Logistic Regression	0.75	0.73	0.82
Logistic Ridge Regression	0.76	0.76	0.83
Random Forest	0.81	0.80	0.88
SVM with RBF Kernel	0.73	0.72	0.84
ANN	0.77	0.76	0.84
CNN	0.76	0.74	0.84

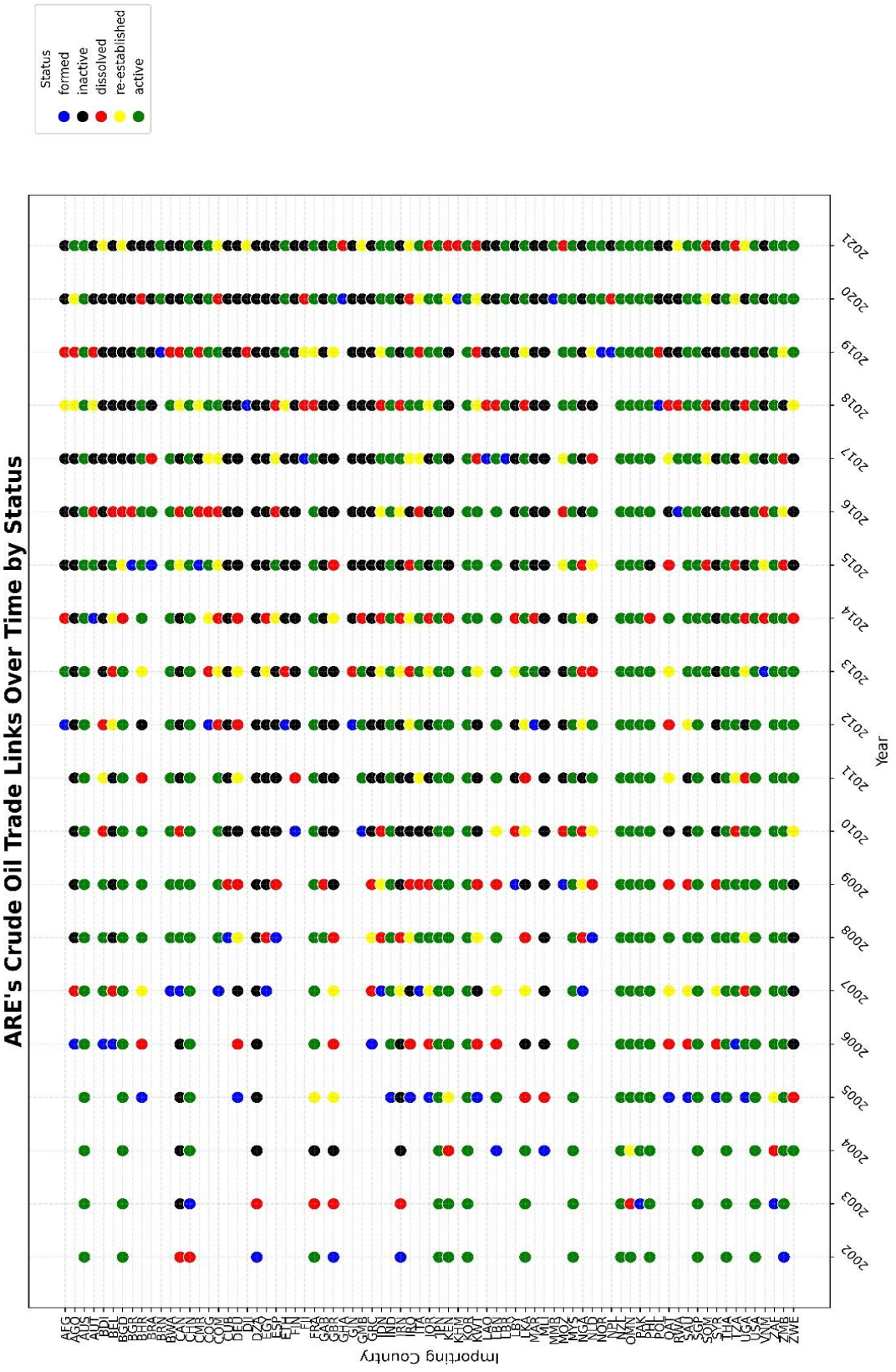


Fig. A.14. Temporal trade link status tags for the United Arab Emirates's trade networks

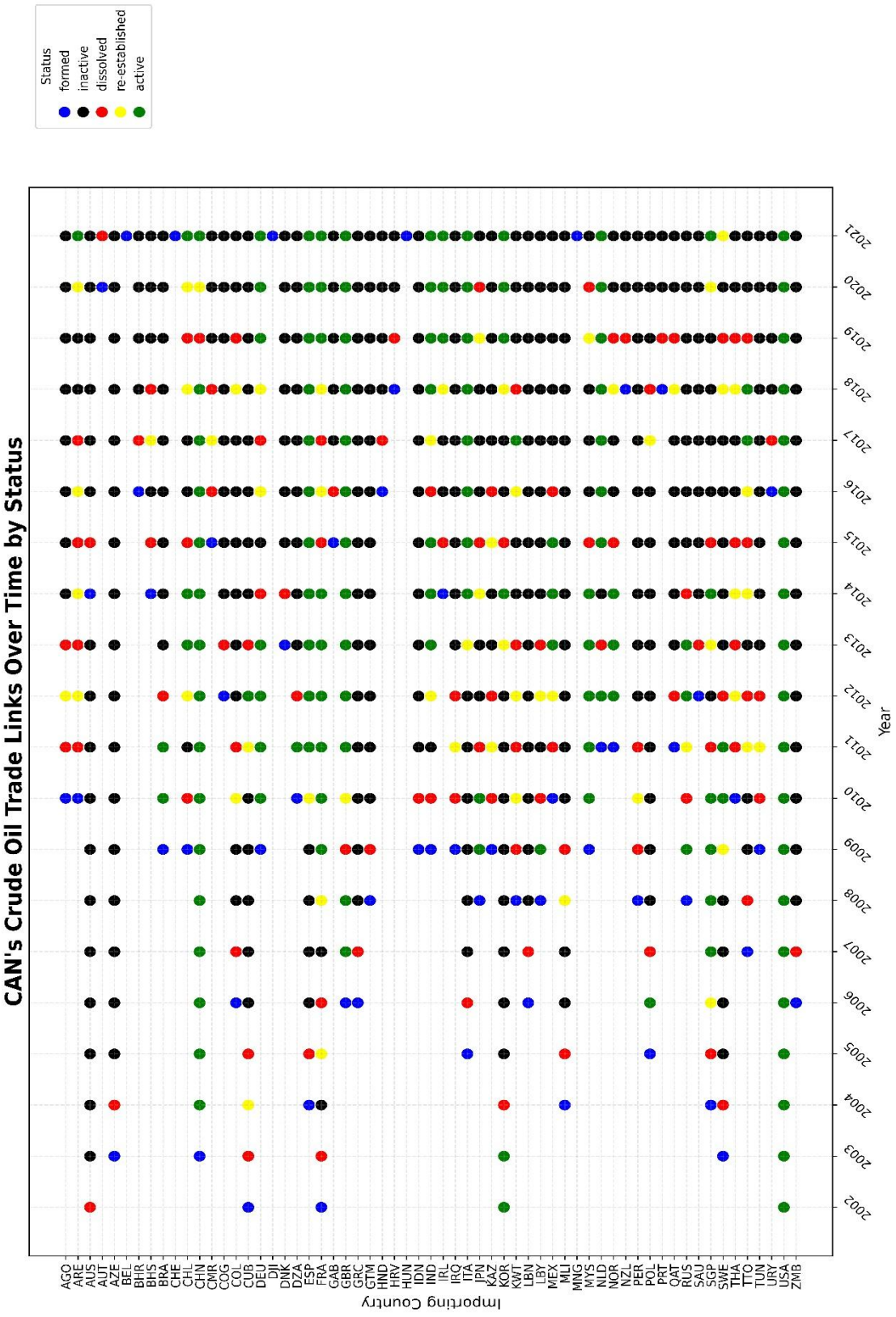


Fig. A.15. Temporal trade link status tags for Canada's trade networks

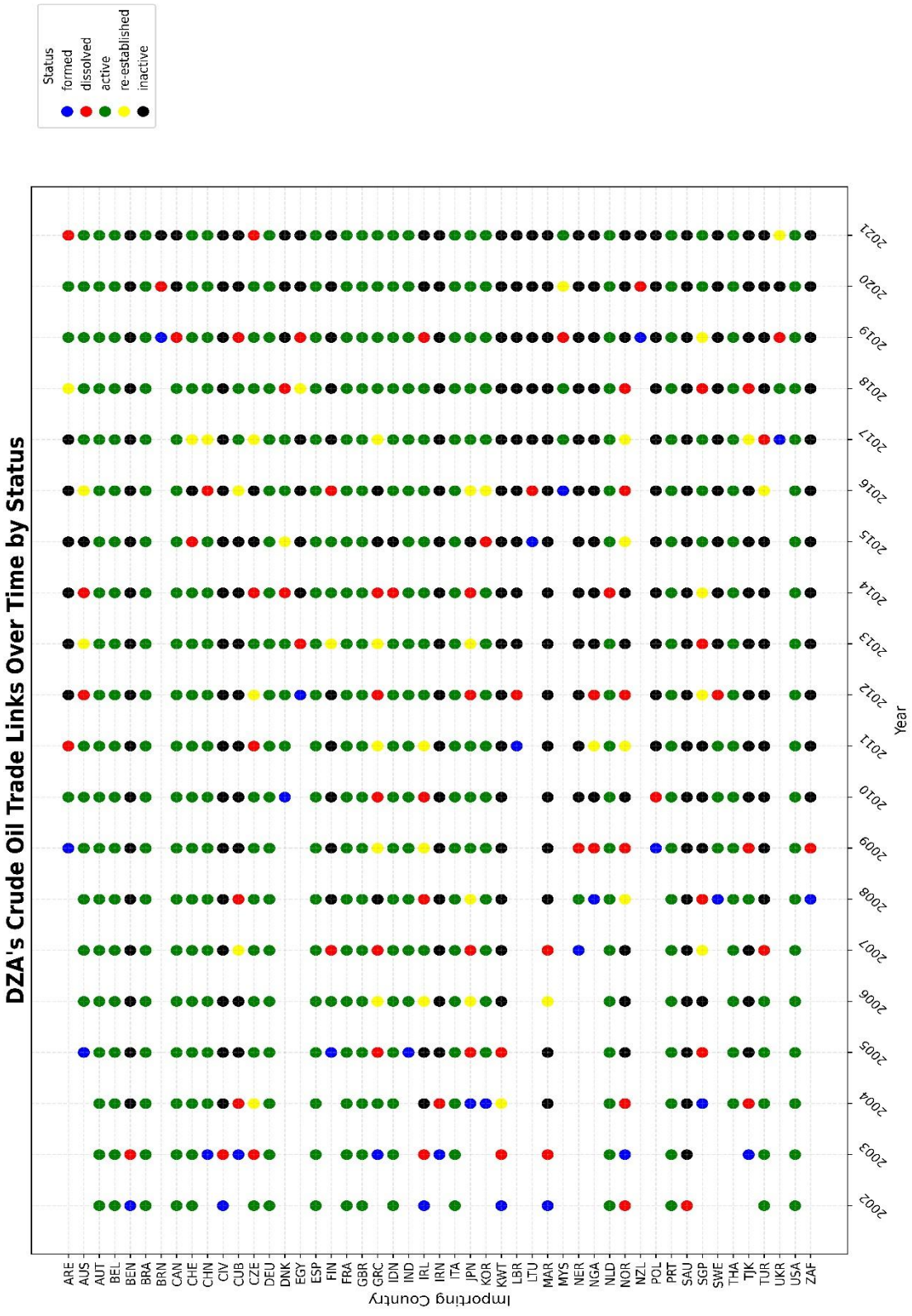


Fig. A.16. Temporal trade link status tags for Algeria's trade networks

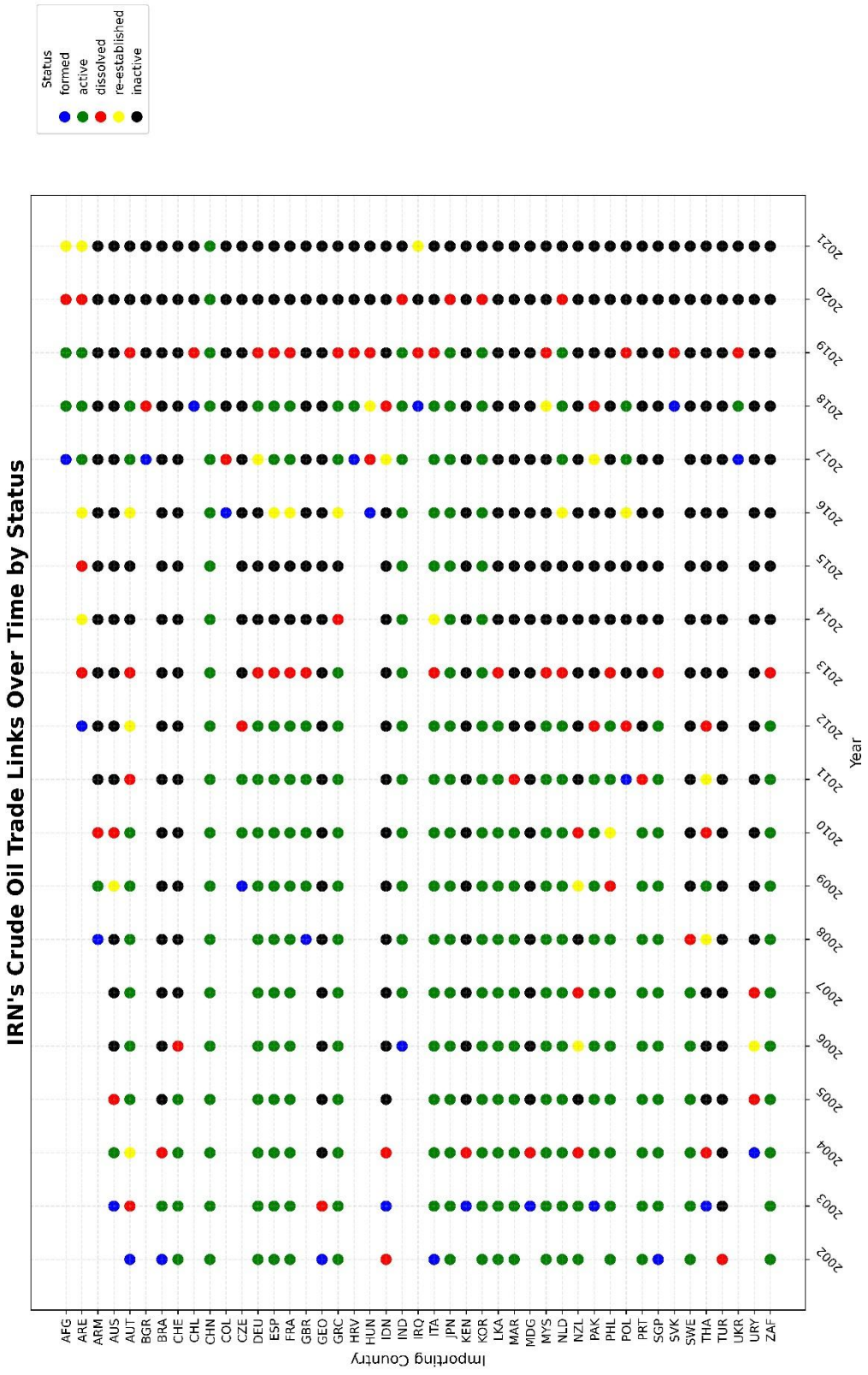


Fig. A.17. Temporal trade link status tags for Iran's trade networks

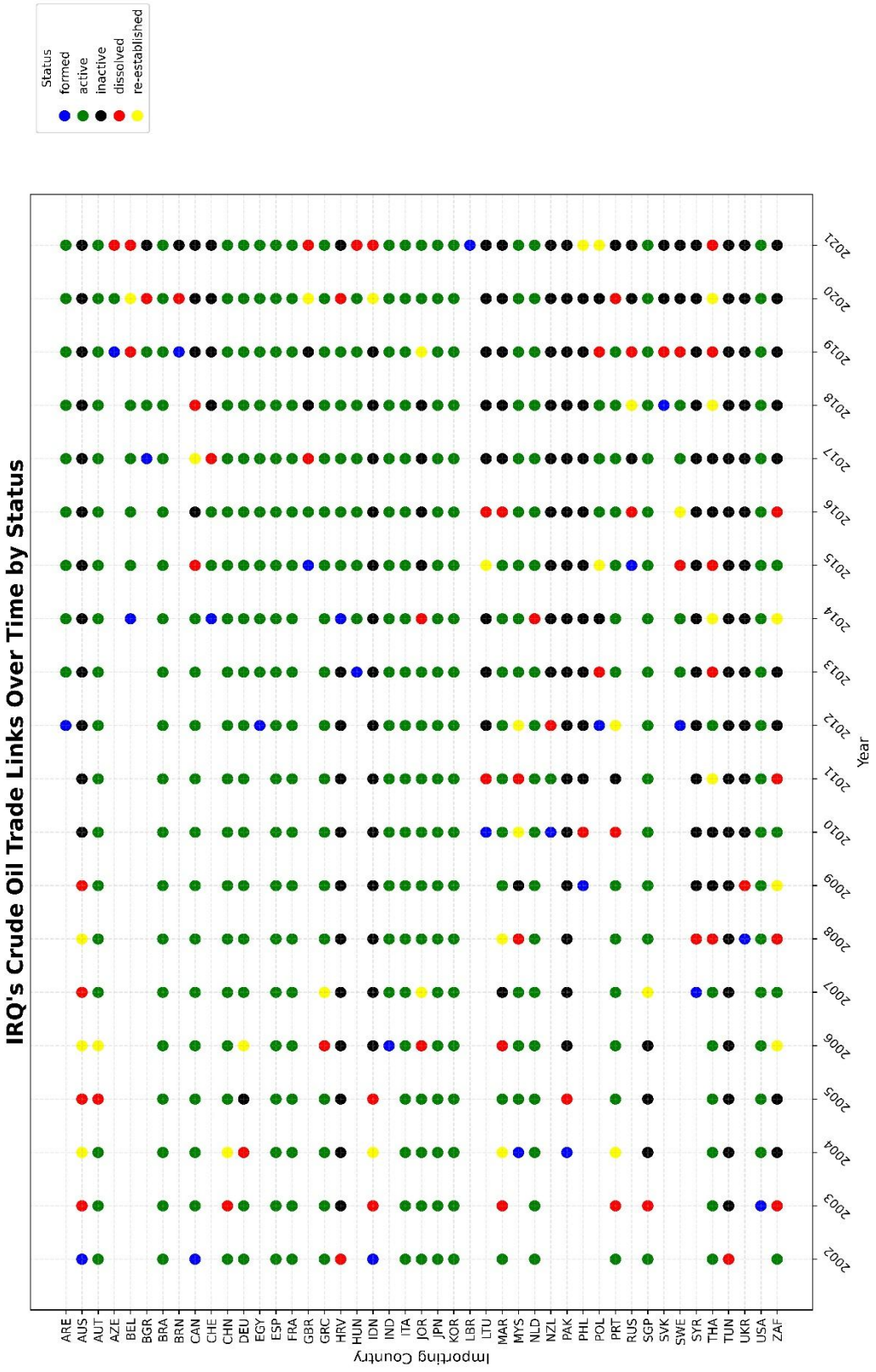


Fig. A.18. Temporal trade link status tags for Iraq's trade networks

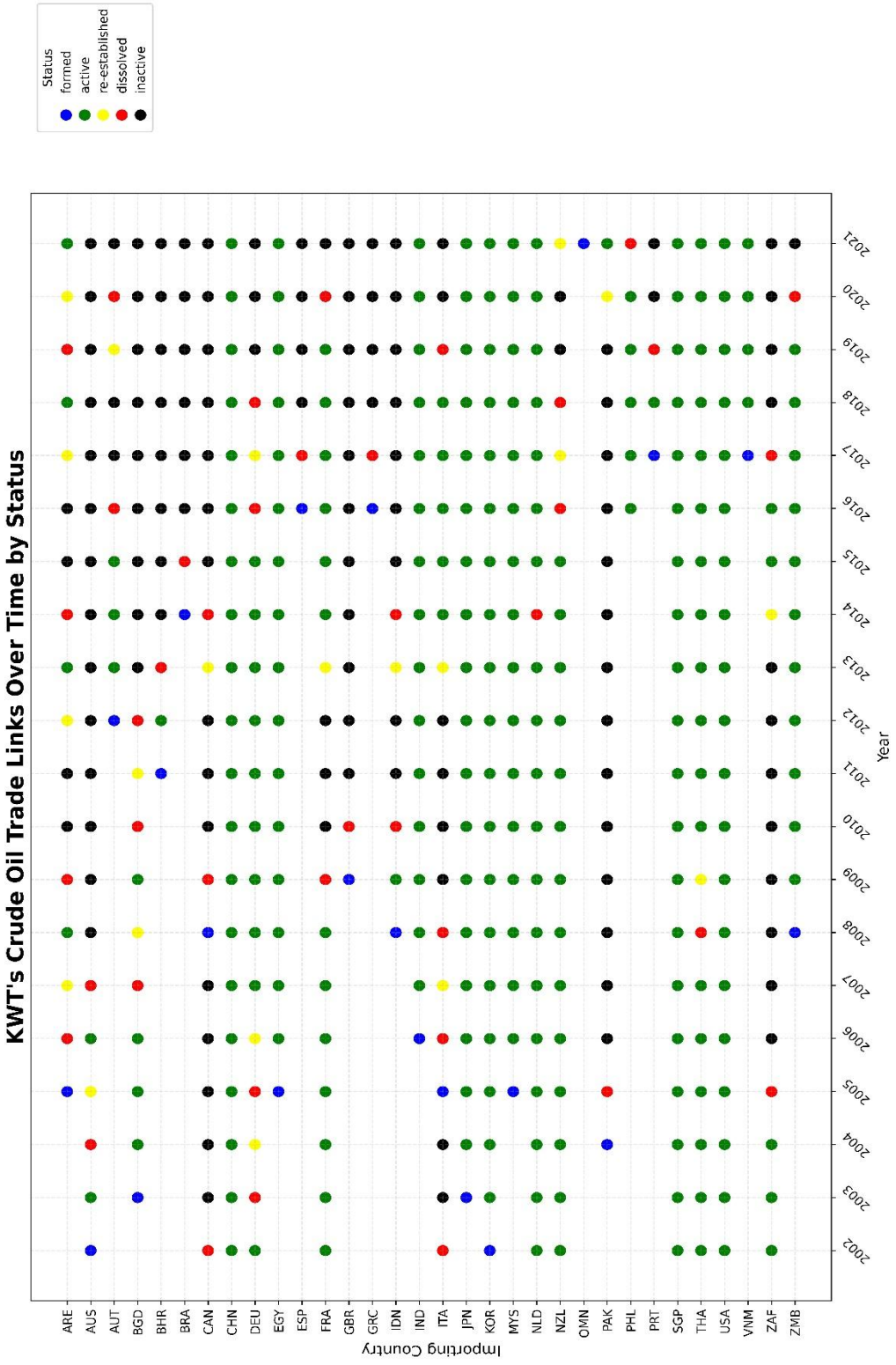


Fig. A.19. Temporal trade link status tags for Kuwait's trade networks

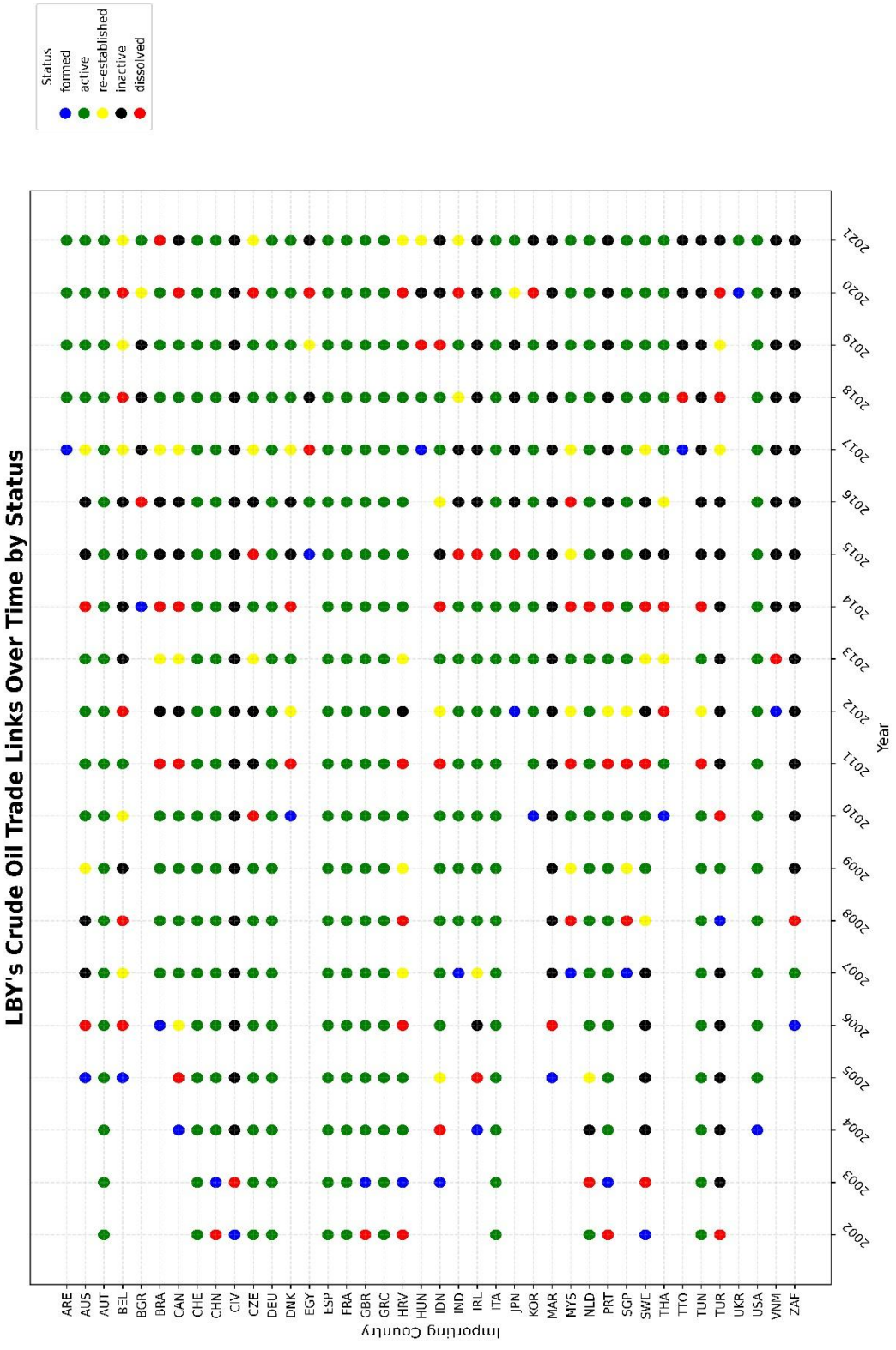


Fig. A.20. Temporal trade link status tags for Libya's trade networks

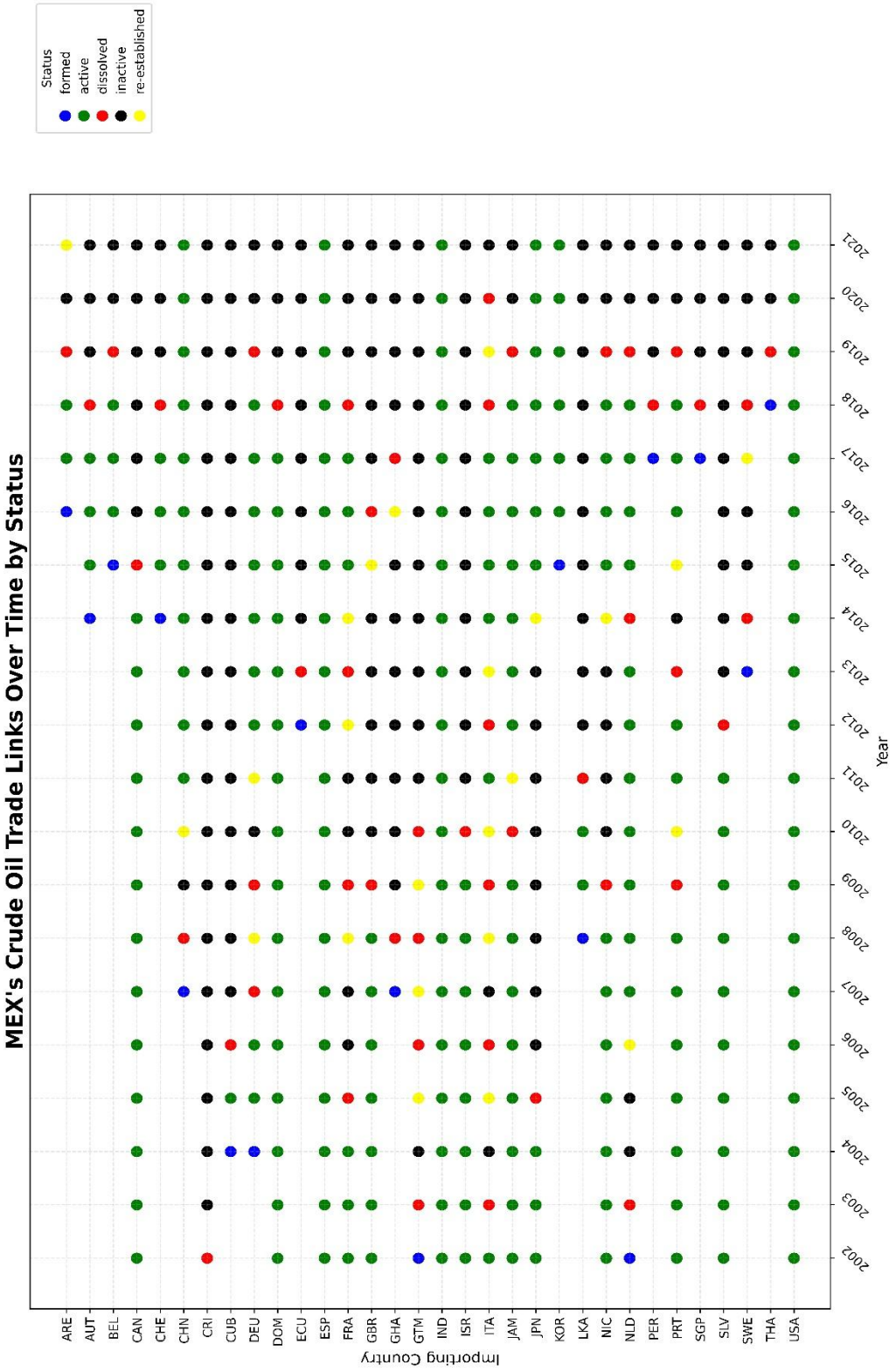


Fig. A.21. Temporal trade link status tags for Mexico's trade networks

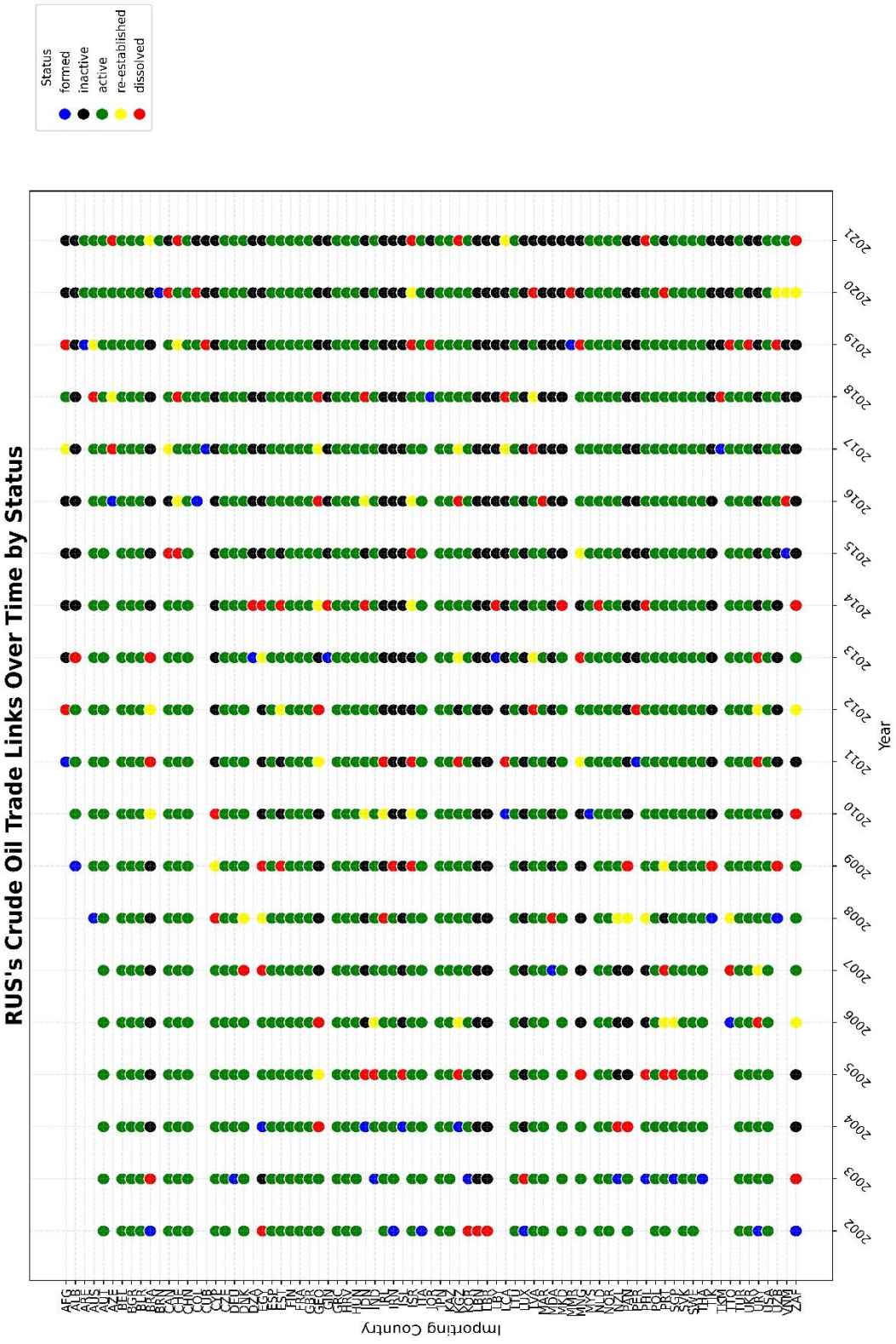


Fig. A.22. Temporal trade link status tags for Russia's trade networks

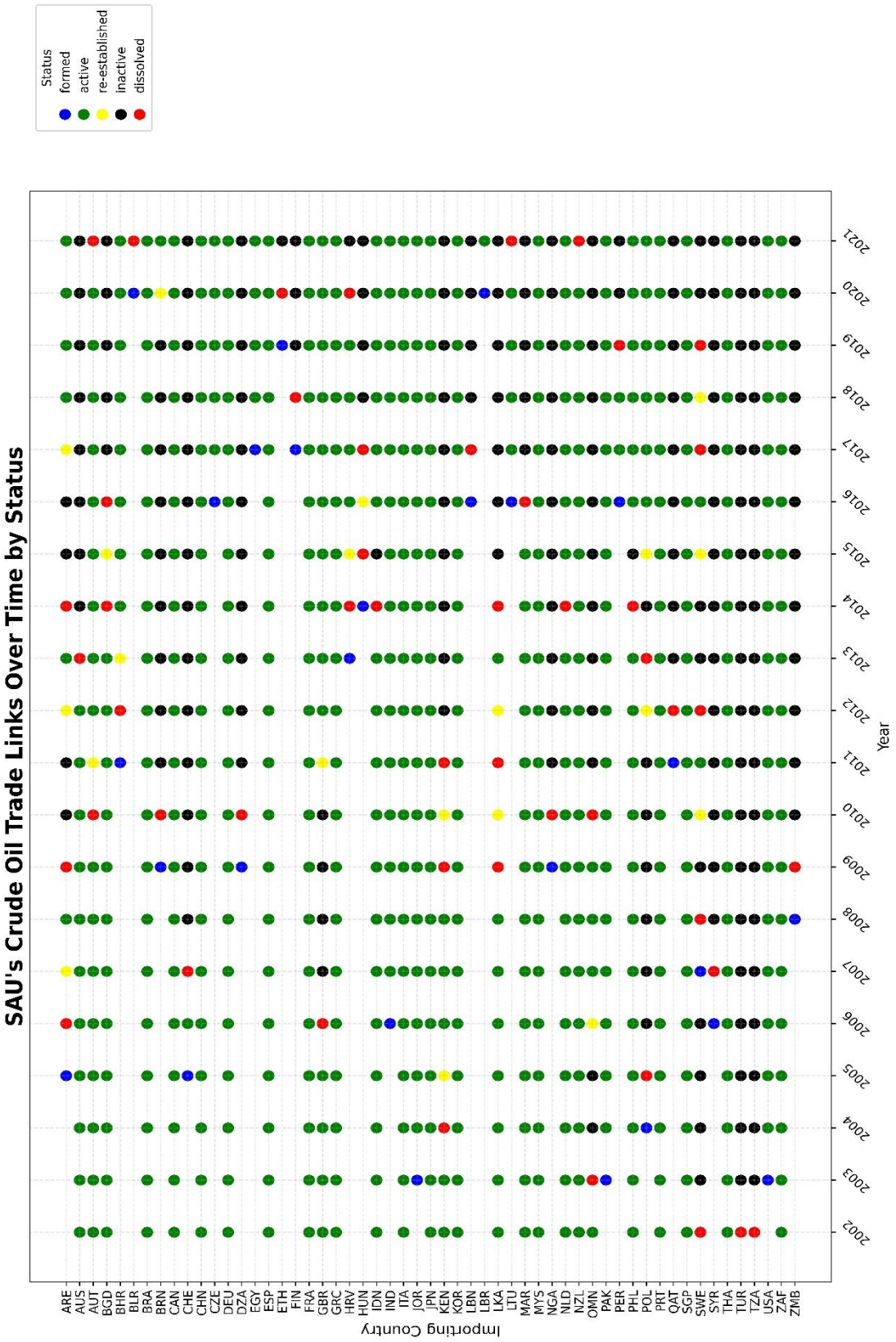


Fig. A.23. Temporal trade link status tags for Saudi Arabia's trade networks

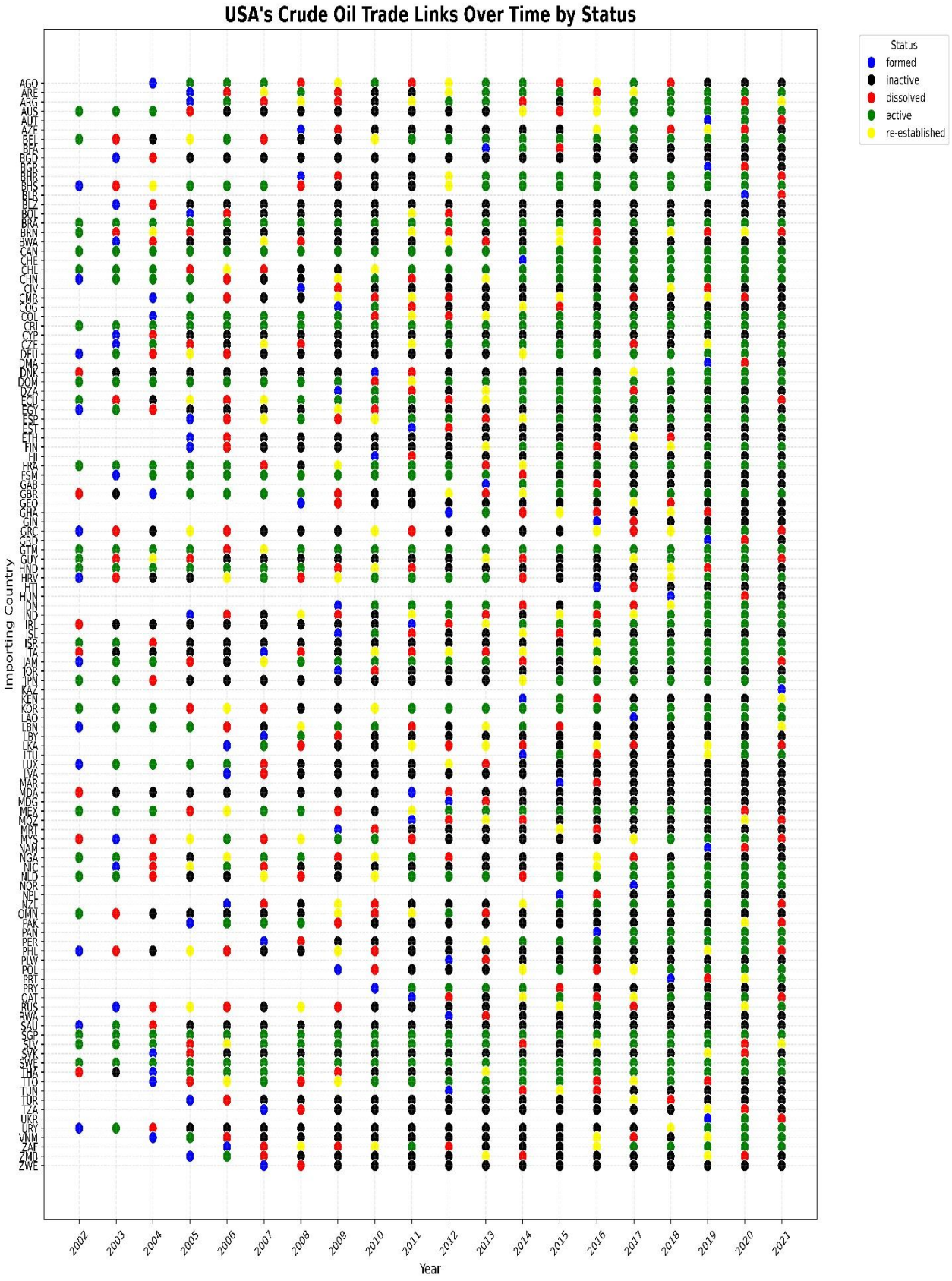


Fig. A.24. Temporal trade link status tags for America's trade networks

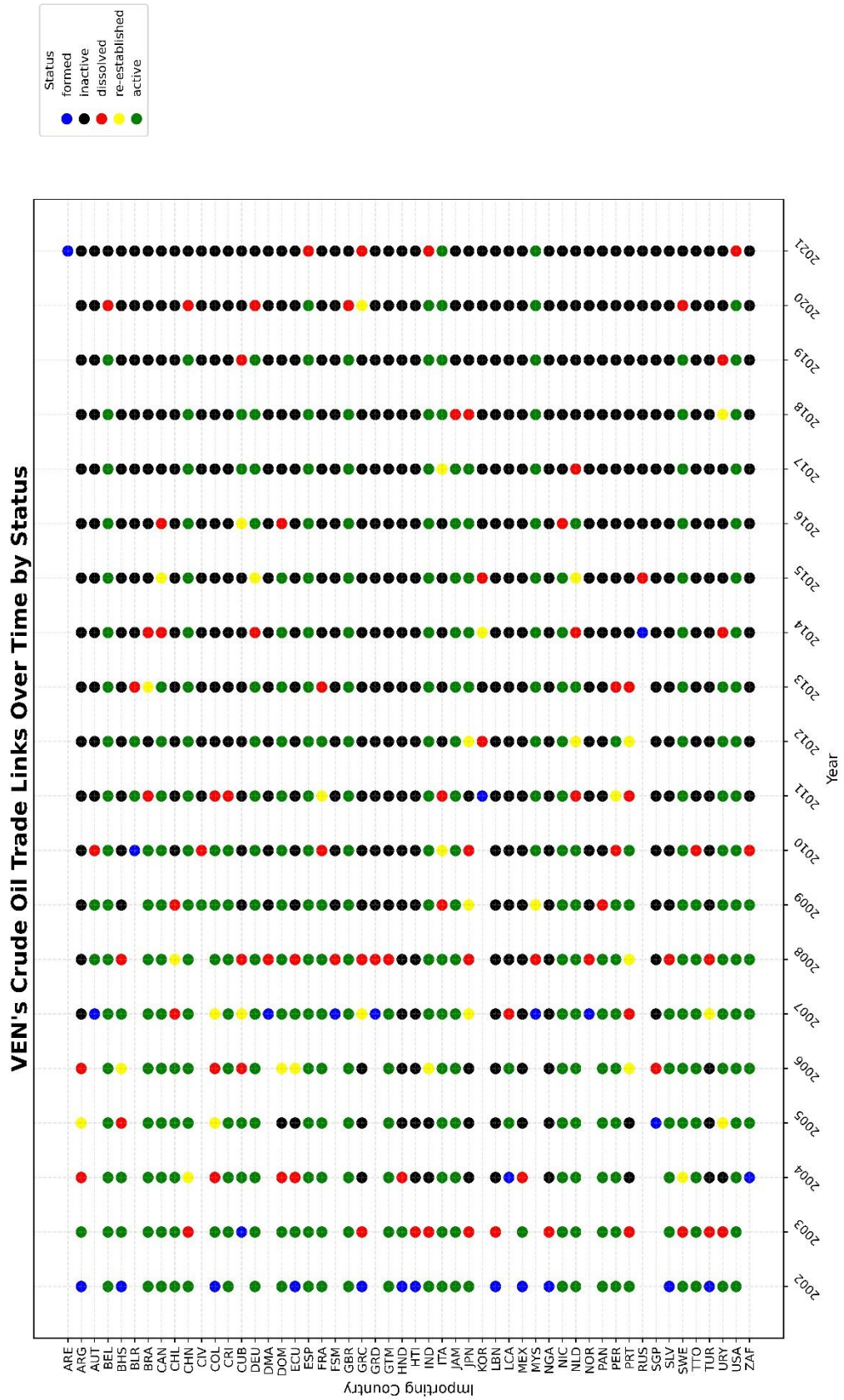


Fig. A.25. Temporal trade link status tags for Venezuela's trade networks

A.8. Detailed analysis of network-based metrics

Fig. 5 demonstrates the year-wise link turnover rates of various oil-exporting countries. From 2003-2006, Canada showed a sustainably high turnover rate, signaling frequent alternations in trade partners, which aligns with the increased integration of Canada in the global energy market during the early 2000s, particularly via the development of tar sands and extended pipeline infrastructure. Moreover, the Iraq War in 2003 and supply disruption in Venezuela seem to have caused a reshuffling in the global crude oil supply chains by incentivizing countries to seek new markets. Iran shows an unprecedented peak in the turnover ratio in 2020, indicating a massive reshuffling of its crude oil export links.

The reasons behind this trend seem to be US sanctions, which severely restricted Iran's capability to export oil legally; on the other hand, oil-importing countries had to comply with US sanctions or seek waivers, leading to sharp alterations in importing preferences. Post 2010, the USA experienced a gradual decline in link turnover rates, which is possibly linked to the shale revolution, which reduced the USA's dependence on oil-exporters, thus providing it with greater foreign policy manoeuvrability. Venezuela demonstrates a low turnover initially, with a late spike in 2021, suggesting a drastic shift owing to economic crises, hyperinflation, and US sanctions. However, in 2021, restrictions were eased in a very limited way, and oil was rerouted to countries willing to bypass US sanctions, such as China and Iran.

The consistent turnover rates of Mexico and Russia depict stable trade relationships. Russia maintained long-term oil trade linkages with European countries such as Germany, Poland, and others, and began to diversify towards Asia only after the 2010s. Similarly, historically, Mexico exports a large share of oil to the US, hence indicating lower partner volatility. However, there is a spike in the turnover rate of Mexico in 2019, demonstrating a shift in energy policy under President Andrés Manuel López Obrador, who came to power in 2018 and reversed energy liberalization reforms, along with emphasizing energy sovereignty and prioritizing domestic refining.

Saudi Arabia, Iraq, and Kuwait have relatively low and stable turnover rates, suggesting stable bilateral agreements, long-term contracts (particularly with Asian countries), and tentatively limited room for diversification due to infrastructural and geopolitical constraints. There has been a surge across multiple countries around 2003-2008 reflecting global reconfiguration post the Iraq War, rise in Chinese demand, and liberalization of global oil markets. Additionally, renewed volatility can be seen in 2020-2021 due to COVID-19, characterized by a drastic drop in global demand, storage constraints, price wars, and policy shifts associated with fossil fuel diplomacy.

Fig. 6 shows the year-wise link substitution index for various oil-exporting countries. Canada portrays a notable spike in the substitution index in the years 2008 and 2021, which means that the traditional buyers of Canada prominently substituted Canadian oil with other exporters. The peak seen in 2007 may reflect early concerns over the environmental impact of oil sands and shifting refinery preferences, while 2021 mostly relates to the post-COVID market realignment and policy alterations under the Biden administration, especially the cancellation of the Keystone XL pipeline, combined with a broader push for clean energy. Iran demonstrates significant peaks in the substitution index in the years 2002, 2003, and 2017.

The time period around 2002-2003 relates to the US invasion of Iraq, regional instability, and enhanced nuclear scrutiny on Iran that led to the importing countries diversifying away from Iran. On the other hand, the peak around 2017 captures the effect of US sanctions under Donald Trump's "maximum pressure" policy, resulting in the replacement of Iranian oil with Saudi, Iraqi, and Russian barrels although this period saw relative easing of sanctions given the fact that the Joint Comprehensive Plan of Action (JCPOA) was signed in 2015 portraying a paradoxical moment. For Libya, the substitution peaks in 2004 and 2017, reflecting elevated volatility due to civil unrest. The year 2004 marks the re-entry of Libya into the global energy market after Gaddafi's decision to abandon WMDs in 2003.

The continuing Libyan Civil War, attacks on oil terminals, and fragmentation of national oil control caused frequent supply disruptions, leading to significant substitution in 2017. Mexico saw a spike in the substitution index in 2002, 2015, and 2017. During the early 2000s, there was a long-term decline in Mexico's oil production from one of its largest oil fields, the Cantrell field. Further, anticipating the reduced reliability, the US, which is Mexico's primary buyer, began diversifying its import sources, in turn switching to Venezuela, Canada, and Saudi Arabia. Moreover, during the 2015-2017 period, the US shale oil boom led to a direct competition between the US's light sweet crude and Mexico's isthmus crude.

Also, the decline in Mexico's oil production around 2017 severely constrained Pemex's ability to fulfil long-term contracts, amplified by the deregulation and policy uncertainty prevailing in the country. There was also a demand-side diversification in the sense that countries like India and China somewhat replaced traditional suppliers like Mexico with oil imports from Iraq, Russia, and Iran. Although Saudi Arabia is mostly considered a "core supplier" in the global oil market, vulnerabilities can be seen by the substitution index of the country peaking in 2003, 2005, and 2016. While 2003 depicts the adjustments in the aftermath of 9/11, the Iraq War, and security concerns in the Gulf, the sharp surge in oil prices in 2004, waves of terrorist attacks, and concerns about Saudi Arabia's capacity to sustainably ramp up output resulted in diversification of supply sources by importers to hedge risk "(Russia, Angola, Brazil and Canada as substitutes) that ultimately led to a higher substitution index in 2005.

The fallout of the 2014-2016 oil price crash also seems to have affected the linkage dynamics of the country. During the price crash, instead of reducing output to stabilize prices, Saudi Arabia resorted to flooding the market with the aim of gaining market share and undercutting higher-cost producers like the US and emerging African and Latin American suppliers. However, this approach backfired as it intensified market volatility and oversupply, leading to some countries diversifying away from Saudi Arabia. Moreover, in 2015, the US lifted the 40-year ban on crude oil export, and Iran aggressively regained market share following the JCPOA agreement of January 2016.

Venezuela's substitution index is, in general, low. The reasons relate to the collapse of the country's crude oil production over the two decades due to multilateral and bilateral sanctions, chronic underinvestment, and mismanagement at PDVSA (Petróleos de Venezuela, S.A.). The majority of the buyers substituted Venezuelan oil in a short window, reflecting no significant year-on-year fluctuation afterward. Also, post-sanctions Venezuela exported crude via clandestine routes or non-transparent mechanisms, which were infrequent, opaque, and

highly erratic. Moreover, the substitution was one-directional and structural because Venezuela exports extra-heavy crude, requiring specialized refineries and is not readily substitutable with light sweet crude.

The minor spike in the substitution index in 2007 likely demonstrates growing concerns regarding the increasing control over PDVSA and the drive for resource nationalism under President Hugo Chávez. Factors like predictable trade relationships, high-quality crude, and policy-driven export strategy culminated in the relative stability of the substitution index of the US from 2002 to 2021, and the relationships are marked by high buyer retention and long-term engagements.

Fig. 7 depicts the year-wise link reactivation rates of various crude oil exporting countries. The years 2004 and 2005 mark the beginning of the UAE's trade link reactivation with a notable spike in 2007, driven by rapid economic growth in China and India, the country's diversification strategy of securing long-term supply contracts, including Asia-Pacific and East Africa, and supply uncertainties from other producers like Iraq and Nigeria. The reactivation rate reached its peak in 2013 owing to post-Arab Spring rebalancing, pivot to Asia, and the operationalization of the Habshan–Fujairah pipeline with the aim of bypassing the Strait of Hormuz.

Canada shows two notable phases of reactivation, the first one during 2004-2006, and the second one between 2010-2012. The first phase corresponds with the rise of Canadian oil sands production, while the second phase is synonymous with infrastructure growth and favourable price differentials owing to the widening gap between WTI and Brent crude, making Canadian oil relatively cheaper for global refiners. Additionally, a spike in 2018 relates to the temporary role of Canada as a substitute for sanctioned oil exporters such as Iran and Venezuela. However, the reactivation rates declined from 2019 onwards, reflecting pipeline constraints, reduced upstream investments, and shifting narratives related to climate policies.

For Algeria, the reactivation rate rises in 2004 and 2006, signifying a global demand surge and the reactivation of supply contracts by the European refineries, particularly in France, Spain, and Italy. Further, the political stability of the country under President Abdelaziz Bouteflika also aided in rebuilding trust with former partners. A sustained and moderate reactivation was seen from 2007 to 2011 due to increased investment in hydrocarbon infrastructure, the US and European concerns over the supply networks of Middle Eastern oil exporters, and Algeria's political neutrality. Given that Algeria proved as a safe haven in North Africa concerning Arab Spring aftershocks, the reactivation rate increased in 2013.

However, a gradual decline and stabilization is seen during 2014-2021, demonstrating crude quality mismatches, pipeline protests, strategic shift towards natural gas, and demand suppression due to COVID-19. The trajectory of Iran underscores the deeply intertwined nature of crude oil with sanctions compliance dynamics, diplomatic breakthroughs, and shifting geopolitics. The peaks and troughs of the reactivation rates of Iran sync closely with the global diplomatic cycle. The early peak in 2004 depicts robust market-driven integration driven by strong global crude oil demand.

A declining but relatively persistent reactivation rate is seen from 2006-2012, emphasizing intensified UN and the US sanctions on Iran's nuclear program, and embargoes on Iranian

Central Bank transactions, although waivers and barter mechanisms, and grey and semi-official channels sustained trade links to a limited extent. A spike in 2016 suggests the enactment of the JCPOA in July 2015 and the subsequent implementation in January 2016, leading to the lifting of nuclear-related sanctions and allowing Iran to re-enter European and Asian markets. However, as the US withdrew from the JCPOA in May 2018, sanctions were reimposed under Donald Trump's "maximum pressure" campaign, which manifested in declining reactivation rates during 2018-2021.

With the toppling of Saddam Hussein's regime due to the US invasion in 2003, there was an oil market reopening in Iraq, leading to the reactivation rate peaking in the same year. Prior to 2003, oil exports of Iraq were heavily constrained under the UN Oil-for-Food Program. The period 2004-2006 demonstrates post-conflict re-engagement with the improvement in export infrastructure of the country and steady production improvements reflecting a cautious but consistent approach. The reactivation rate declined in subsequent years with occasional peaks owing to growing market competition, security concerns, budgetary instability, and infrastructure bottlenecks. Occasional peaks reflect sporadic opportunistic sales or short-term contracts rather than sustained policy shifts.

A strategic ordinance of trade engagement aligned with global demand shifts and regional stability is being revealed by the reactivation rates of Kuwait. Reflecting Kuwait's integration into the global crude oil market following a decade of post-Gulf War uncertainty, the reactivation rate peaked in 2004. Amid OPEC+ realignments and mercurial market preferences driven by sanctions on other producers, there is a moderate resurgence in 2017. Finally, declining rates from 2018 onwards suggest the transition of the country from re-engagement towards long-term trade consolidation.

The critical reliance of oil-market reengagement on institutional credibility and political stability is aptly depicted through the trajectory of Libya. The 2005 to 2009 wave of reactivation marked the end of Western sanctions, thus allowing Libya to reintegrate into the global economy. With the fall of Muammar Gaddafi and NATO intervention, the country sought to reinstitute oil production and exports, which is reflected in the spike in the reactivation rates in 2012 and 2013. However, the enthusiasm was short-lived and Libya descended into renewed civil conflict in 2014, characterised by competing governments in Tripoli and Tobruk along with ongoing attacks on oil terminals by militias, which restricted the reactivation capacity in later years.

Moreover, the sharp peak in 2017 exhibits partial stabilization of the country with the reopening of major export terminals, enhanced autonomy of the National Oil Corporation, and exemption from OPEC+ production cuts. Turning to Mexico, the reactivation rates demonstrate two distinct spikes, one in 2005 and another in 2008, corresponding to enhanced export capacity and global price shock-driven diversification, respectively. With the energy reforms in 2013, the reactivation rate again peaked in 2014, followed by a diminishing trend owing to competition from shale oil and long-term export commitments.

As a result of the Orange Revolution in Ukraine in 2004-2005, Russia's relations with the EU were strained, and thus, it started diversifying its oil trade portfolio. To ensure connectivity to South Korea, Japan, and China, investments in the East Siberia–Pacific Ocean (ESPO) pipeline were accelerated. Simultaneously, with the breakup of the Yukos oil company, state control was consolidated under Rosneft. All these events culminated in the spike in

reactivation rates during 2005-2008. From 2009 onwards, Russia pursued dual-track energy diplomacy in which it kept legacy buyers in Europe, along with focusing on Asia and the Middle East, showing sustained reactivation.

A low but somewhat stable reactivation rate during 2013-2021 shows that although Western sanctions were imposed on Russia following the annexation of Crimea in 2014, it managed to solidify trade ties through pre-existing long-term contracts and the completion of the ESPO pipeline. Given the uncertainty associated with the US invasion of Iraq in 2003, Saudi Arabia, being a swing producer, stepped in to fill supply gaps, leading to a jump in the country's reactivation rate in 2005. On account of the surge in global demand in 2006 and 2007, expansion in Asian refining infrastructure, and intermittent disruptions faced by Iraq, Nigeria, Venezuela, and Iran, Saudi Arabia became an attractive substitute as an oil-exporting country.

Again, the reactivation rate peaked in 2010, followed by elevated levels in 2011 and 2012, marked by the recovery of oil demand after the global financial crisis, Saudi Arabia's role as a stabilizing force in OPEC, and the impact of Fukushima disaster in Japan in 2011 that triggered a shift away from nuclear power. Further, although the reactivation rate peaked for a short while in 2015, it consistently declined afterwards owing to Saudi Arabia's market share defense strategy, in which, after the price collapse during early 2015, it increased the production levels instead of cutting them. But the declining reactivation rates afterwards point to long-term contract saturation, internal policy shifts related to Vision 2030, and energy transition dynamics.

The US demonstrates consistently high reactivation rates from 2004 to 2016, followed by a gradual decline. The reasons behind this trend relate to the post-Iraq war demand boom, significant exports of refined petroleum products (jet fuel, gasoline, diesel), and the commodity supercycle of the mid-2000s. From 2008 onwards, the enhanced production of shale oil from the Permian, Bakken, and Eagle Ford basins contributed to the higher reactivation rates. The legislative shift pertaining to the repealing of the 40-year-old ban on crude oil exports in 2015 immediately opened trade with Europe, Asia, Latin America, and Africa.

Subsequently, there was a saturation of core trade networks, a trade war with China during 2018-19, and COVID-19 that led to a gradual decline in reactivation rates of the US. Venezuela's reactivation index started rising in 2004, reached its peak in 2006, remained higher in 2007, and then declined gradually. The national oil strike in Venezuela in 2002-03 paralyzed PDVSA, but from 2004 onwards, it focused on re-establishing dissolved relationships, which coincided with the global commodity supercycle, Further, Chávez used oil strategically as a diplomatic tool, in turn offering preferential deals to allies.

Although, high reactivation rate was sustained in 2007 but early signs of fragility were visible through prevailing underinvestment in upstream assets, politicization of PDVSA, and the resulting diversification of oil revenues towards domestic populist spending. Consequently, there was a gradual decline in the reactivation rate from 2008 onwards owing to the global financial crisis, production inefficiency, and an increasing share of exports tied with debt repayments (specifically under China's oil-for-loan program).

Fig. 8 represents the year-wise link persistence ratio of various crude oil exporting countries. The most persistent trade relationships are those of Saudi Arabia, where the persistence ratio

remains high till 2013 with only a modest decline afterwards. This stability is synonymous with long-term supply contracts, minimal disruption in trade linkages even during oil price collapses or OPEC decisions, and corroborates the strategic position of the country as a swing producer and OPEC leader. For Russia, the persistence ratio remained high till 2013 with a slight decline afterwards.

This trend tantamounts to diversification of export routes, imposition of Western sanctions post Crimea annexation in 2014, and the gradual shift from spot market exports to pipeline-linked bilateral deals. Venezuela's persistence ratio was higher till 2009 but showed a sharp decline afterwards, reaching a near collapse by 2021. The drivers behind this scenario are the prevalence of political instability through radicalization of oil policy, sanctions imposed by the US in 2017 and 2019, and the collapse of PDVSA's operational capacity attributed to mismanagement and corruption.

The US shows a low persistence ratio initially, with a gradual increase post-2014, given that the US was primarily an oil importer in the very beginning, with exports limited to only refined products. This trend changed post-2014 due to the shale oil revolution and the lifting of the export ban. However, trade volatility with Europe and China, particularly during the Trump era, prevented stable link formation. Iran's persistence ratio was initially very high but started falling since 2013. The reasons for the collapse include EU embargo and US led international sanctions, and severed long-term contracts.

The persistence ratio jumped slightly during 2017-18 but again fell as Trump withdrew from the JCPOA in 2018. Although Mexico started with a very high persistence ratio, it started falling gradually since 2007, with higher values in 2011-12 and 2015-2017. The period of 2002-2006 concerns peak production from Mexico's Cantarell field, with exports primarily to the US and only a limited amount to Europe and Asia. From 2007 to 2010, Cantarell's output began to plummet along with increased competition from the US's shale oil. Mexico opted for strategic diversification during 2011-2012, but further production decline and policy uncertainty under AMLO's government led to a sharp fall in the persistence ratio in later years.

Canada has consistently low values of persistence ratio because it exports primarily to one buyer, which is the US, and most inactive links remained inactive owing to a lack of infrastructure. Through the early 2000s, the UAE maintained a relatively high persistence ratio due to stable trade networks anchored in long-term bilateral contracts with Asian importers. This stability persisted in later years but declined after 2014, mirroring a shift in the export policy of Abu Dhabi National Oil Company (ADNOC) towards market-responsive trading mechanisms in place of rigid contracts. This was further corroborated by rising global competition leading to a decline in trade link continuity.

During 2002-2013, Algeria maintained a high persistence ratio owing to regional proximity to European importers and stable demand based on a volume-based contract model. The slight drop in the ratio from 2014-2016 was due to the global oil price crash in mid-2014, and rising competition from other exporters, combined with domestic fluctuations. However, later years show stabilization and recovery through increased spot market activity and reengagement with Asian importers. Libya demonstrates a non-linear but geopolitically significant trajectory in terms of persistence ratio. The period from 2002 to 2010 marks Libya's

reintegration into the global market following years of sanctions and Gaddafi's diplomatic normalization with the West.

However, given the Arab Spring and NATO-backed uprising, Gaddafi's regime collapsed in 2011, and Libya descended into civil war. This manifested in the collapse of oil exports, leading to a complete fragmentation of Libya's export continuity. There was partial recovery during 2018-2019 due to limited political stabilization, but later years saw renewed decline due to renewed military offensives, blockade by Haftar-aligned forces in early 2020, and COVID-19.