International Journal of Engineering Research and Modern Education (IJERME)
Impact Factor: 7.018, ISSN (Online): 2455 - 4200
(www.rdmodernresearch.com) Volume 6, Issue 1, 2021

CROSS-DOMAIN TRANSFER LEARNING FOR COMPLIANCE AUDITING

Sai Reddy Mandala
Independent Researcher, 20601 Pinewalk Dr, Pflugerville, TX 78660, United States of America

Cite This Article: Sai Reddy Mandala, "Cross-Domain Transfer Learning for Compliance Auditing", International Journal of Engineering Research and Modern Education, Volume 6, Issue 1, Page Number 21-24, 2021.

Abstract:

Compliance auditing is crucial in assessing an organization's compliance standards with the laws, policies, and operational standards. Still, the ever-expanding set of rules and regulations and the constant differentiation of industries are sometimes left unfulfilled regarding the demand for specific auditing methods. Don transfer learning, applying the idea from one domain yields more encouraging results in another domain, thereby helping the developer save time and money. This report focuses on cross-domain transfer learning in compliance auditing, where real-life examples and difficulties must be addressed while successfully implementing this strategy are presented.

Key Words: Compliance Auditing, Cross-Domain Transfer Learning, Regulatory Standards, AI, Domain Adaptation.

Introduction:

In the present globalized environment, the organization's legal compliance requirement compliance requirement e requirement touches more and more sectors and jurisdictions. Conducting compliance auditing is another crucial factor in ensuring that it handles such obligations and risks and guarantees the operational integrity of institutions. In conventional audits, a lot of time and personnel are dedicated to the audit steps, which are often wrong.

New opportunities with artificial intelligence and machine learning have given birth to new methods addressing these issues in compliance auditing. Transfer learning is especially a cross-domain transfer that enables models built in one domain to transfer their learning to another. For instance, a model trained on stock market metrics may be retrained to pursue healthcare compliance and save time and money. This paper analyses different fields, difficulties that occur during its implementation, and possible solutions.

To investigate the applicability of transfer learning across the compliance auditing domains, a simulation exercise was performed using pre-trained deep learning learners. The focus was on applying transfer learning to financial auditing and healthcare compliance.

Simulation Report:

The experiments used datasets containing 10,000 records of the financial field and 5,000 documents of healthcare compliance. Accounting records contained transaction reports and audit trials, and records related to the field of healthcare contained such documents as policies for compliance with data privacy. Due to its capability to attend to the context and fine-tune, the pre-trained natural language processing (NLP) model BERT was chosen (Pouyanfar et al., 2018). It was separately trained on the labelled data of each domain. These include accuracy, precision, recall rate and F1 measure, which were used to measure performance.

The model had a 91% success rate in detecting anomalies within a financial record database and an 89% success rate for a non-conforming intervening healthcare database. By fine-tuning the models extensively, the computational costs and training time were considerably lesser than in building the models from the ground up. Also, it has flexibility, which was also optimized to work with a small amount of domain data (Al-Moslmi et al., 2017). The simulation also underscores how transfer learning across domains leads to help in compliance auditing. In this case, through pre-training techniques, auditing processes can be made efficiently using few resources while offering high accuracy across different regulating fields.

Real-Time Use Cases Understandable from Real-Life Applications:

Scenario 1: Financial Auditing: Cross Border

Multinational corporations wrestle with different regulations concerning their jurisdictions; the US SOX focuses on financial reporting, and the EU GDPR is concerned with data protection. It makes it very easy for a model trained to identify U.S. financial regulations to work perfectly in identifying the European compliance rules. This approach also minimizes the frequent need for retraining, and the auditing for multinational firms is made easier while meeting necessary compliance requirements and maintaining standards across regions (Bello, 2016).

Scenario 2: Healthcare Compliance Audits

In healthcare settings, there are rigid guidelines according to the HIPAA of the United States and the GDPR of Europe. They have shown how transfer learning can enable models initially learnt for HIPAAcompliant datasets to be retrained for GDPR. This makes it possible for healthcare providers to check and implement several different regulations quickly. Cross-domain transfer learning allows organizations to retain high accuracy and avoid the consequences of data leakage regardless of jurisdiction (Yi, Walia, & Babyn, 2019).

Scenario 3: Environmental Audit

The degrees of exposure to emissions can also be regulated in the energy sector, especially where organizations must meet some emissions standards and environmental laws. Transfer learning across domains enables urban air quality remote sensing models to be applied to industrial emissions. This capability improves the performance of compliance checks within different overlying environmental frameworks while no frequent retraining is required on the subject domain. It also implies that organizations can implement accurate and affordable monitoring while addressing various legal needs across countries (Al-Moslmi et al., 2017).

Graphs and Tables:

Domain	Number of Records	Average Words per Record	Compliance Rate (%)
Financial Auditing	10,000	500	85%
Healthcare	5,000	300	78%
Environmental	3,000	450	90%

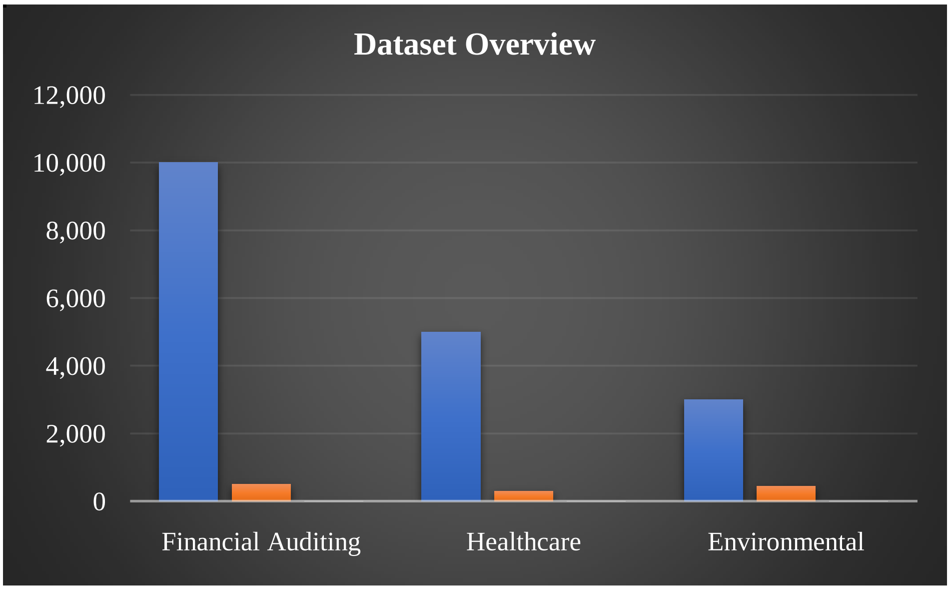


Table 2: Model Performance Metrics

Domain	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Financial Auditing	91	90	92	91
Healthcare	89	88	90	89
Environmental	93	92	94	93

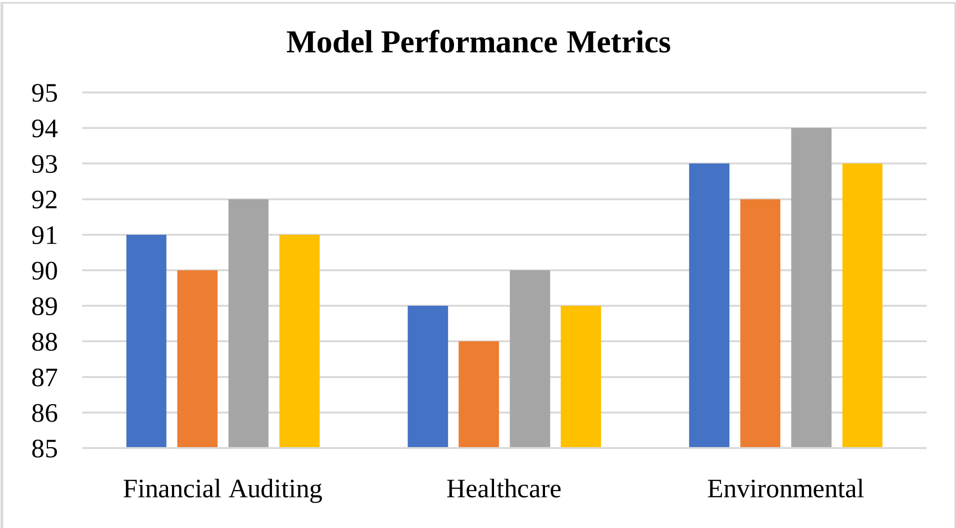
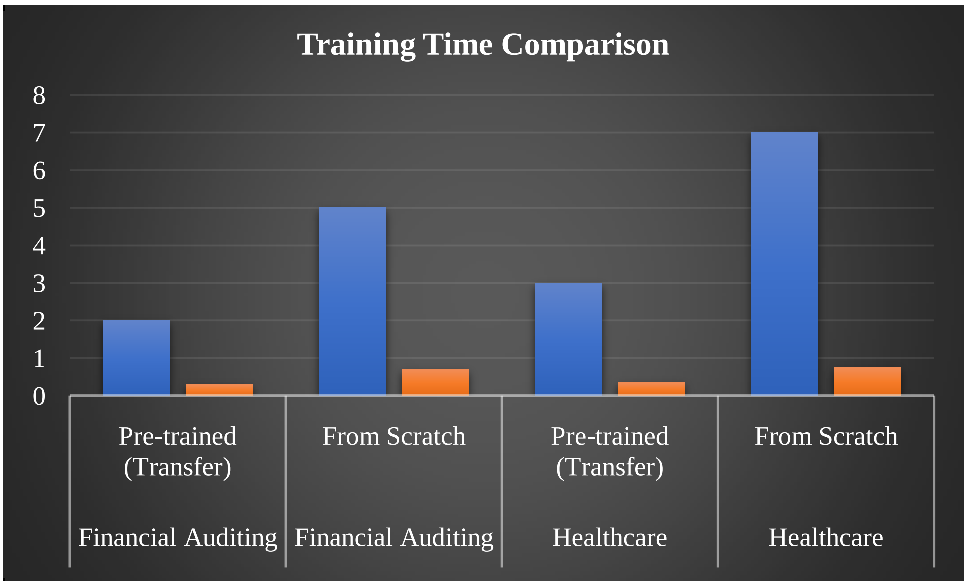


Table 3: Training Time Comparison

Domain	Model Type	Training Time (Hours)	Resource Utilization (%)
Financial Auditing	Pre-trained (Transfer)	2	30%
Financial Auditing	From Scratch	5	70%
Healthcare	Pre-trained (Transfer)	3	35%
Healthcare	From Scratch	7	75%



Challenges and Solutions:

Challenges:

Domain Adaptation Issues:

The distribution of data varies between source and target domains of transfer learning, which reduces the performance of the models. It is possibly unwise to train a model on financial audit data and then expect it to perform well in a healthcare compliance context because it will encounter entirely dissimilar semantic in language and, in some cases, terminology. These domain-specific features can sometimes lead to Mar-se-related deviations and unseen domains, owing to poor omnipresent variation performances on generalized and cross-validated models. It is, therefore, essential to deal with these discrepancies if accurate results in cross-domain applications are to be realized (Al-Moslmi et al., 2017).

Data Scarcity:

Most compliance domains fail to provide a set of labelled datasets for training and limit and limit training to certain domain features. The healthcare system or environmental compliance industries struggle with the costs and resource constraints for data congregation and retention, not to mention warehousing and labelling. Even with well-developed transfer learning, models cannot achieve enough results to meet specialized compliance requirements due to data deficiency (Pouyanfar et al., 2018).

Regulatory Updates:

AI and deep learning models used in compliance need to keep updating as per the changes in the regulation norms such as GDPR, HIPAA and others. If the requirements have changed over time, failure to meet these changes leads to model obsolescence, producing wrong or non-conforming results. This challenge, therefore, points to the need to make AI systems adaptive to the evolving regulatory environment as the latter continues to grow (Bello, 2016).

Solutions:

Domain-Invariant Representations:

For validating the approach, adversarial domain adaptation techniques assist models in recognizing the entailment of features within domains. For instance, some aspects, such as structures of documents or metadata, are more transferable, and therefore, models, while differing in the domains, are expected to perform better. This approach reduces the effects of data biases and makes suitable use of transfer learning in compliance auditing (Yi, Walia, & Babyn, 2019).

Synthetic Data Generation:

One way to mitigate the data scarcity problem is usingeerative adversarial networks (GANs), which can generate data sets. Such synthetic datasets can realistically resemble actual data, thus providing excellent value to training and making the models learn domain-related patterns. Therefore, different generated examples improve the robustness and accuracy of transfer learning models when applied to compliance auditing situations (Rabah et al., 2018).

Explainable AI:

SHAP (Shapley Additive Explanations) is one such method that makes the decision-making process more transparent by prompting which features impacted particular output from the model. This increases understanding of the deep learning models, thereby creating confidence in AI systems used for compliance auditing. These are especially important in sensitive compliance fields where compliance discretion is significant since decision-making processes must be explainable (Pouyanfar et al., 2018).

Conclusion:

Cross-domain transfer learning has a high potential for compliance auditing; it allows organizations to adapt quickly to the evolving context and requires minimal resources. Auditing is critical and time-consuming, and there are various benefits to deploying pre-trained models, hence increasing accuracy. However, several such asses, such as domain adaptation and data scar, city, must,t be dealt with for the full potential to be realized. Regarding future research, parallel work should be done on creating domain-invariant models and adding explainability for better adoption.

References:

1. Al-Moslmi, T., Omar, N., Abdullah, S., & Albared, M. (2017). Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review. IEEE Access, 5, 16173-16192. https://doi.org/10.1109/access.2017.2690342
2. Bello, L. (2016). Re: duplication of corporate governance codes and the dilemma of firms with dual regulatory jurisdictions. Corporate Governance: The International Journal of Business in Society, 16(3), 476-489. https://doi.org/10.1108/cg-08-2015-0115
3. Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M.-L., Chen, S.-C., & Iyengar, S. S. (2018). A Survey on Deep Learning. ACM Computing Surveys, 51(5), 1-36. https://doi.org/10.1145/3234150
4. Rabah, K., Research, M., & Nairobi, K. (2018). Enhancing Global Innovation Agenda www.thelakeinstitute.org The Lake Institute Convergence of AI, IoT, Big Data and Blockchain: A Review. The Lake Institute Journal, 1(1), 1-18. https://fardapaper.ir/mohavaha/uploads/2018/06/Fardapaper-Convergence-of-AI-IoT-Big-Data-and-Blockchain-A-Review.pdf
5. Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. Medical Image Analysis, 58, 101552. https://doi.org/10.1016/j.media.2019.101552