pharmacists advancing healthcare®

WHEN AI GOES ASTRAY
Understanding Model Drift

Scott D. Nelson, PharmD, MS, FAMIA, ACHIP

The field of medicine is a rapidly evolving landscape, causing significant challenges for machine learning and artificial intelligence (AI). Due to changes in clinical practice, underlying data, and the relationships between variables, the performance of AI models degrade over time – a phenomenon known as "model drift". Model drift not only affect predictive models but also poses challenges for the newer generative AI models, such as large language models (LLMs). Understanding and addressing model drift is crucial for maintaining the accuracy and reliability of AI systems in healthcare.

CHANGES IN THE UNDERLYING DATA

The decrease in model performance is not unique to AI models already in use. As Lasko et al. point out, most models fail to transport between settings or locations due to poor model performance resulting from differences in site-specific clinical practices and inherent differences in data generating processes between sites.1 This is why it is important to retrain and calibrate predictive models on local data. For AI models already in use, changes in clinical practices and data generating processes over time can likewise cause poor model performance. Examples of these changes include data drift, label drift, concept drift, covariate shift, and changes due to external factors or seasonal variation.

TYPE OF DRIFT/CHANGE	DESCRIPTION	EXAMPLES
Data Drift (Feature Drift)	Changes in the statistical properties or patterns of the input data.	▪ Introduction or withdrawal of medications. ▪ Changes in data collection or documentation (e.g. new EHR system). ▪ Shift in patient demographics (e.g. aging patient population) ▪ Adoption of new clinical practice protocols or guidelines.
Label Drift	Changes in the definition of the model output (labels).	▪ Redefined of clinical outcomes (e.g. how re-admissions are counted). ▪ Changes in diagnostic criteria. ▪ Updates in coding systems (e.g. ICD-9 to ICD-10 codes). ▪ Policy changes affecting quality metrics or treatment protocols.
Concept Drift	Changes in the relationship between input features and output labels.	▪ Evolving medical knowledge altering understanding of diseases. ▪ New treatment protocols becoming standard of care. ▪ Technological advancements in diagnostic tools. ▪ Emergence of new diseases or antibiotic-resistant infections.
Covariate Shift	Changes in input data distribution, when the input-output relationship remains the same.	▪ Applying a model trained on adults to pediatric patients ▪ Using the same model in different clinical settings. ▪ Expanding AI use from physicians to nurses and pharmacists altering input data distribution.
External Factors	Changes due to external conditions like policy changes or global events that affect the data.	▪ New healthcare regulations or drug formulary adjustments. ▪ Drug shortages leading to alternative treatments. ▪ Pandemics and epidemics.
Seasonal Variation	Predictable, time-based changes affecting data patterns.	▪ Increases in respiratory illnesses during certain seasons. ▪ Seasonal health campaigns impacting healthcare utilization patterns. ▪ Weather-related health issues (e.g. Heatwaves).

DATA DRIFT

Data drift, also known as feature drift, occurs when there are changes to the statistical properties of the data going into the model (input features). This type of drift can significantly impact the performance of AI models, as they rely on consistent data patterns to make accurate predictions. When the characteristics of the input data change over time, the model's assumptions about the data no longer hold true, leading to decreased performance. Examples include:

1. **Introduction or withdrawal of medications:** The pharmaceutical landscape is dynamic, with new drugs entering the market and others being withdrawn, or even changes in prescribing preferences.
2. **Changes in data collection or documentation:** Variations in how data is collected or documented can also cause data drift. For example, if a hospital updates its electronic health record (EHR) system, the way data is recorded might change, or even its location in the EHR, which the model may then miss completely.
3. **Patient demographics:** Shifts in patient demographics, such as age, gender, or ethnicity, can also lead to data drift. If a model was trained on a specific patient population, changes in the demographics of the patients being treated can alter the input data distribution. One such example includes our aging population in general.
4. **Clinical practice changes:** Updates in clinical guidelines or treatment protocols can change the input data characteristics. For example, if a new treatment protocol is adopted widely, the data reflecting patient outcomes and treatments will change, leading to data drift.

LABEL DRIFT

Label drift occurs when the definition of the model output (label) changes over time. For example, the operational definition for what is considered a readmission could change. This type of drift can significantly impact the performance of AI models, as they rely on consistent definitions of the output to make accurate predictions. When the criteria for what constitutes a particular label change, the model's predictions can become less reliable, especially since they are effectively predicting something else. Examples of label drift include:

1. **Redefinition of clinical outcomes:** In healthcare, the definitions of clinical outcomes can evolve. For instance, the criteria for what constitutes a hospital readmission might change based on new clinical guidelines or policies. If an AI model was trained on an older definition, its predictions may no longer align with the new criteria.
2. **Changes in diagnostic criteria:** Medical conditions and diseases are sometimes redefined as new research emerges. For example, the diagnostic criteria for conditions like diabetes or hypertension can change, affecting the labels used in the model.
3. **Updates in coding systems:** Healthcare systems often update their coding systems, such as transitioning from ICD-9 to ICD-10 codes. These updates can change the labels used in the data, leading to label drift. Models trained on the old coding system may struggle to accurately predict outcomes using the new codes.
4. **Policy changes:** Changes in healthcare policies can also lead to label drift. For example, new regulations might alter the criteria for what is considered a quality metric, such as patient satisfaction scores or adherence to treatment protocols. These changes can affect the labels used in the model.

CONCEPT DRIFT

Concept drift occurs when the relationship between the input features and output labels changes over time. This type of drift can significantly impact the performance of AI models, as they rely on stable relationships between inputs and outputs to make accurate predictions. Examples of concept drift include:

1. **Evolving medical knowledge:** As medical research advances, new discoveries can change the understanding of diseases and their treatments.
2. **Changes in treatment protocols:** Updates in clinical guidelines and treatment protocols can alter the relationships between input features and outcomes. For instance, if a new treatment becomes the standard of care for a condition, the outcomes for patients receiving this treatment may differ from those who received the previous standard of care.
3. **Technological advancements:** The introduction of new medical technologies, such as advanced imaging techniques or novel diagnostic tools, can change the way diseases are detected and treated.
4. **Population health changes:** Shifts in population health trends, such as the emergence of new diseases or changes in the prevalence of existing conditions, can also cause concept drift. For instance, the rise of antibiotic-resistant infections can change the relationship between antibiotic use and patient outcomes to change.

COVARIATE SHIFT

Covariate shift is a type of data drift where the relationship between the input features and output labels remains the same, but the distribution of the input features changes. This shift can occur when the population or environment from which the data is drawn changes, leading to a new distribution of inputs that the model was not trained on. While the underlying relationship between inputs and outputs remains constant, the model's performance can still degrade because it encounters data that is different from what it has seen before. Examples include:

1. **Demographic changes:** If a model was trained on data from a specific demographic group, applying it to a different demographic can lead to covariate shift. For example, a model trained on adult patients may not perform well when applied to pediatric patients, as the distribution of health conditions and responses to treatment can differ significantly between these groups.
2. **New departments or hospitals:** When a new department or hospital starts using a model already in use, the patient population may differ from the one the model was trained on. For instance, a model trained in a tertiary care hospital may not perform as well in a community hospital due to differences in patient demographics, disease prevalence, and treatment protocols. This is important to keep in mind as health systems consolidate and expand.
3. **Changes in user base for generative AI:** In the context of generative AI, covariate shift can occur when the user base changes. For instance, if a model initially rolled out to physicians is later introduced to nurses and pharmacists, the types of questions and needs of these different user groups can vary, leading to changes in prompts and a shift in the input data distribution.

EXTERNAL FACTORS

External factors can cause drift when there are changes due to external conditions, such as policy changes or global events. Examples include:

4. **Policy changes:** Healthcare policies are subject to change, and these changes can affect the data used by AI models. For example, new regulations might alter the criteria for patient eligibility for certain treatments or services. Additionally, changes to drug formularies can significantly change ordering and treatment patterns.
5. **Drug Shortages:** Drug shortages can impact the availability of medications, leading to changes in treatment patterns. For instance, if a commonly used drug becomes unavailable, healthcare providers might switch to alternative treatments.
6. **Pandemics and Epidemics:** Global health events, such as the COVID-19 pandemic, can cause significant shifts in healthcare data. During a pandemic, the types of patients seen, the treatments provided, and the outcomes observed can all change dramatically. For example, the COVID-19 pandemic broke algorithms for reading chest x-rays, which then needed to be retrained. Then when we were better able to treat COVID-19, the models all broke again!

SEASONAL VARIATION

Seasonal variation refers to predictable changes that occur at certain times of the year. While some seasonal changes can be anticipated, they can still cause problems for model performance if not properly accounted for. Examples include:

1. **Influenza and RSV seasons:** The onset of influenza and respiratory syncytial virus (RSV) seasons can lead to an increase in respiratory illnesses. This seasonal surge can change the distribution of input features, such as the number of patients presenting with respiratory symptoms.
2. **Seasonal health campaigns:** Public health campaigns, such as flu vaccination drives, can change patient behavior and healthcare utilization patterns.
3. **Weather-related health issues:** Extreme weather conditions, such as heatwaves or cold snaps, can lead to an increase in weather-related health issues. For instance, heatwaves can cause a rise in heat-related illnesses, while cold snaps can lead to more cases of hypothermia and frostbite.

ADDRESSING MODEL DRIFT

In order to maintain the accuracy and reliability of AI systems in health care, it is essential to continuously monitor and update models. In model development, different model training methods exhibit varying rates of drift. For example, neural networks may drift more slowly compared to regression models.[2] Additionally, collaborating with clinical experts and leadership is crucial to understand and anticipate changes in practice, data, or policies that could impact model performance. This proactive approach helps in designing models that are more resilient to drift.

CONTINUOUS MONITORING

The cornerstone of effectively addressing model drift is continuous monitoring of model performance, a process known as "algorithmovigilance".[3] Algorithmovigilance involves the ongoing monitoring and evaluation of healthcare algorithms, akin to pharmacovigilance for medications. The goal is to continuously assess the safety, efficacy, and performance of AI models used in healthcare. For predictive models, we can monitor the statistical properties of the features going into the model, predictive performance, user interactions, and the model's impact on clinical outcomes. For generative AI models, we can monitor the number of tokens or queries per day, the average response time, user satisfaction, the semantic similarity of the queries, the ROUGE score (semantic similarity between the prompt and the output), and other standard performance metrics where appropriate.

When a model starts to drift or reaches a threshold requiring action, it may need to be recalibrated, replaced, or even decommissioned. This ensures that the AI system remains effective and reliable in the dynamic healthcare environment.

CONCLUSION

In the rapidly evolving field of health care, maintaining the accuracy and reliability of AI models is a continuous challenge due to model drift. By understanding the different forms of drift—data drift, label drift, concept drift, covariate shift, external factors, and seasonal variation—healthcare professionals can better anticipate and address these changes. Implementing strategies such as continuous monitoring, regular retraining, and collaboration with clinical experts ensures that AI systems remain effective and reliable tools. As AI continues to play an increasingly vital role in healthcare, proactive management of model drift will be essential for delivering high-quality patient care and improving clinical outcomes.

REFERENCES

1. Lasko TA, Strobl EV and Stead WW. Why do probabilistic clinical models fail to transport between sites. *NPJ Digit Med.* 2024; 7(1): 53. PMID:38429353
2. Davis SE, Lasko TA, Chen G et al. Calibration drift in regression and machine learning models for acute kidney injury. *J Am Med Inform Assoc.* 2017; 24(6): 1052-61. PMID:28379439
3. Embi PJ. Algorithmovigilance-advancing methods to analyze and monitor artificial intelligence-driven health care for effectiveness and equity. *JAMA Netw Open.* 2021; 4(4): e214622. PMID:33856479

Updated: January 2025