Hewlett Packard Enterprise

HPE Machine Learning Data Management Software for Healthcare & Life Sciences

Accelerate insights from ML with data pipelines and autoscaling


Investment in healthcare and life sciences continues to rise, driven by an aging population and the promise of new drugs and breakthrough technologies. The rapid development, creation, and rollout of the new COVID vaccines and the promise of RSV vaccines continues to energize this market. Machine learning has played a role in developing these breakthroughs and will do so for future advances. However, there are some unique challenges to developing ML models for healthcare and life sciences:

- **Data is unstructured:** Most healthcare data isn't stored in a database or files but in physical charts, EMRs, X-Rays, MRIs, audio files, and even DNA sequences.
- **Data is disparate and heterogenous:** Data is in different formats (text, images, video, and audio) and spread across different systems from providers to payors.
- **Data sets are large:** Most use cases have petabytes of data and millions of records that need to be continually processed to derive accurate results.
- **Reproducibility is key:** Organizations need to reproduce any outcome by identifying what data was used and what models were used to produce what results.
- **Data changes more frequently than the ML model:** The ML model is relatively static in comparison to the volumes of data being changed and updated.
- **Not enough experimentation:** The time required for running and re-running projects can be prohibitive, which means less experimentation and smaller datasets.

Patient record

X-rays

Genomics

Modernized care

Improved diagnosis

Genomic sequencing

Optimized treatment

Efficient clinical trials

How HPE Machine Learning Data Management Software can help

HPE Machine Learning Data Management Software empowers data engineering teams to automate complex data pipelines. Our unique architecture is cost-effective at scale and enables sophisticated data transformations across any type of data. We provide auto-scaling and parallelized processing of multi-stage, language-agnostic pipelines with data versioning and data lineage tracking.

- **Support for any type of data:** Work with structured data such as EMRs, test results, claims data, and unstructured data such as PET, CAT, MRI scans, X-rays, audio recordings, and ultrasounds.
- **Aggregates data** from multiple formats and multiple sources into datasets for ready analysis.
- **Scales easily to petabytes** of data and millions of records using autoscaling and parallel processing, enabling frequent model iterations and faster data processing.
- **Captures data lineage by versioning data, pipelines, and code** so that any outcome can be reproduced and audited.
- **Automatically processes data changes** rather than process old data (incremental data processing) so that models are using the most current information and reducing compute costs.
- **Supports parallel and concurrent processing**, reducing the time required to run data pipelines and iterate on ML models, helping to converge on a champion model faster.

Top use cases for HPE Machine Learning Data Management Software

Modernizing care infrastructure

Processing millions of patient records, applying OCR and natural language processing to derive meaning from unstructured data and clinician audio recordings form the basis of modernizing patient records. With this modernization, the records can then be used by the data scientist teams to provide ML models that improve patient diagnosis, patient treatment plans, and better outcomes.

Improve diagnostics

Consider bone density scans which are often used to diagnose or assess the risk of osteoporosis. The problem with current manual processes is they are error-prone and inconsistent depending on the radiologist who is interpreting the results. New systems based on ML image processing are replacing these old systems resulting in better osteoporosis detection, reduced burnout of radiologists, fewer false positives, and a better healthcare experience.

Genomic sequencing

A liquid biopsy can identify infectious diseases or biomarkers of pathogens freeing clinicians from the diagnostic maze so they can return patients back to health quickly and safely. ML models help clinicians avoid invasive, low-yield, and sequential diagnostic tests that can delay treatment for the most vulnerable hospitalized patients.

Optimized patient treatment

Before enhanced image processing and ML models, the existing processes to determine tumor size and volume relied on manual measurements by the oncologist, which is error-prone and inconsistent. New systems use ML models that can create volumetric calculations based on ultrasound images and determine cancer growth projections. Using this data and the historical success of different drugs is paving the way for a more targeted and successful approach to treating the patient with just the right amount of chemotherapy at the right time.

Risk analysis and assessment of clinical trials

Clinical trials are often lengthy, and there is a myriad of information that needs to be captured and analyzed—patient records, clinician recordings, lab results, x-rays, MRIs, etc. There are significant costs and potential life-saving opportunities in stopping a trial if the projections are unfavorable. With ML models, data science teams can compare the results of prior trials, current trial results, and timing, and can more accurately predict the outcome of a specific trial. Armed with this information, they can add members to the trial or cancel it. The goal is to get to market faster with less risk.

a00131537ENW, Rev. 7