

Additional file 5 to: Can it be checked? A cross-sectional meta-research audit of access to, and reporting in, machine learning studies of pharmaceutical manufacturing and quality control.
Phani Kumar Balagam.

Search strategy

All searches ran on 23 September 2026 (UTC times below) through each source's public API, from `harvest.js` (SHA-256 4d03e1255af25c2670a6248faa8cbc6fc1a674cb76dd207d1f951c8b4355115a). To re-run: `node harvest-run.mjs OUT_DIR` (Additional file 6). Years 2018-2026 were harvested; the first-online window 2019-2025 was applied afterwards.

PubMed (E-utilities `esearch`, field tags `[tiab]`)

Started 2026-09-23T13:58:04.390Z; finished 2026-09-23T13:58:09.918Z; count reported by the source 567; records returned 567.

```
("machine learning"[tiab] OR "deep learning"[tiab] OR "neural network"[tiab] OR "neural networks"[tiab] OR "random forest"[tiab] OR "support vector machine"[tiab] OR "support vector regression"[tiab] OR "gradient boosting"[tiab] OR "XGBoost"[tiab] OR "artificial intelligence"[tiab] OR "decision tree"[tiab] OR "k-nearest neighbor"[tiab] OR "gaussian process regression"[tiab] OR "ensemble learning"[tiab]) AND ("pharmaceutical manufacturing"[tiab] OR "drug manufacturing"[tiab] OR "biopharmaceutical manufacturing"[tiab] OR "biomanufacturing"[tiab] OR "bioprocess"[tiab] OR "bioprocessing"[tiab] OR "bioreactor"[tiab] OR "process analytical technology"[tiab] OR "continuous manufacturing"[tiab] OR "pharmaceutical process"[tiab] OR "pharmaceutical quality"[tiab] OR "quality by design"[tiab] OR "critical quality attribute"[tiab] OR "critical quality attributes"[tiab] OR "critical process parameter"[tiab] OR "critical process parameters"[tiab] OR "granulation"[tiab] OR "tableting"[tiab] OR "tablet press"[tiab] OR "tablet compression"[tiab] OR "tablet coating"[tiab] OR "direct compression"[tiab] OR "roller compaction"[tiab] OR "blend uniformity"[tiab] OR "content uniformity"[tiab] OR "spray drying"[tiab] OR "hot melt extrusion"[tiab] OR "freeze drying"[tiab] OR "lyophilization"[tiab] OR "crystallization"[tiab] OR "batch release"[tiab] OR "real-time release"[tiab] OR "release testing"[tiab] OR "stability
```

testing"[tiab] OR "shelf
 life"[tiab] OR "visual inspection"[tiab] OR "quality control"[tiab] OR
 "process monitoring
 "[tiab] OR "fault detection"[tiab] OR "dissolution"[tiab]) AND
 ("pharmaceutical"[tiab] OR
 "pharmaceuticals"[tiab] OR "drug product"[tiab] OR "drug
 products"[tiab] OR "drug substanc
 e"[tiab] OR "active pharmaceutical ingredient"[tiab] OR
 "biopharmaceutical"[tiab] OR "biop
 harmaceuticals"[tiab] OR "biologics"[tiab] OR "monoclonal
 antibody"[tiab] OR "monoclonal a
 ntibodies"[tiab] OR "vaccine"[tiab] OR "vaccines"[tiab] OR
 "therapeutic protein"[tiab] OR
 "tableting"[tiab] OR "tablet press"[tiab] OR "tablet
 compression"[tiab] OR "tablet coating
 "[tiab] OR "direct compression"[tiab] OR "roller compaction"[tiab] OR
 "blend uniformity"[t
 iab] OR "content uniformity"[tiab] OR "process analytical
 technology"[tiab]) AND 2018:2026
 [dp] AND journal article[pt] NOT review[pt]

Europe PMC (REST search, TITLE_ABS)

Started 2026-09-23T13:59:43.341Z; finished 2026-09-23T13:59:46.346Z; count reported by the
 source 499; records returned 499.

(TITLE_ABS:"machine learning" OR TITLE_ABS:"deep learning" OR
 TITLE_ABS:"neural network" O
 R TITLE_ABS:"neural networks" OR TITLE_ABS:"random forest" OR
 TITLE_ABS:"support vector ma
 chine" OR TITLE_ABS:"support vector regression" OR TITLE_ABS:"gradient
 boosting" OR TITLE_
 ABS:"XGBoost" OR TITLE_ABS:"artificial intelligence" OR
 TITLE_ABS:"decision tree" OR TITLE_
 _ABS:"k-nearest neighbor" OR TITLE_ABS:"gaussian process regression"
 OR TITLE_ABS:"ensembl
 e learning") AND (TITLE_ABS:"pharmaceutical manufacturing" OR
 TITLE_ABS:"drug manufacturin
 g" OR TITLE_ABS:"biopharmaceutical manufacturing" OR
 TITLE_ABS:"biomanufacturing" OR TITLE_
 _ABS:"bioprocess" OR TITLE_ABS:"bioprocessing" OR
 TITLE_ABS:"bioreactor" OR TITLE_ABS:"pro
 cess analytical technology" OR TITLE_ABS:"continuous manufacturing" OR
 TITLE_ABS:"pharmace
 utical process" OR TITLE_ABS:"pharmaceutical quality" OR
 TITLE_ABS:"quality by design" OR
 TITLE_ABS:"critical quality attribute" OR TITLE_ABS:"critical quality
 attributes" OR TITLE_
 _ABS:"critical process parameter" OR TITLE_ABS:"critical process
 parameters" OR TITLE_ABS:

"granulation" OR TITLE_ABS:"tableting" OR TITLE_ABS:"tablet press" OR
 TITLE_ABS:"tablet co
 mpression" OR TITLE_ABS:"tablet coating" OR TITLE_ABS:"direct
 compression" OR TITLE_ABS:"r
 oller compaction" OR TITLE_ABS:"blend uniformity" OR
 TITLE_ABS:"content uniformity" OR TIT
 LE_ABS:"spray drying" OR TITLE_ABS:"hot melt extrusion" OR
 TITLE_ABS:"freeze drying" OR TIT
 TLE_ABS:"lyophilization" OR TITLE_ABS:"crystallization" OR
 TITLE_ABS:"batch release" OR TIT
 TLE_ABS:"real-time release" OR TITLE_ABS:"release testing" OR
 TITLE_ABS:"stability testing
 " OR TITLE_ABS:"shelf life" OR TITLE_ABS:"visual inspection" OR
 TITLE_ABS:"quality control
 " OR TITLE_ABS:"process monitoring" OR TITLE_ABS:"fault detection" OR
 TITLE_ABS:"dissoluti
 on") AND (TITLE_ABS:"pharmaceutical" OR TITLE_ABS:"pharmaceuticals" OR
 TITLE_ABS:"drug pro
 duct" OR TITLE_ABS:"drug products" OR TITLE_ABS:"drug substance" OR
 TITLE_ABS:"active phar
 maceutical ingredient" OR TITLE_ABS:"biopharmaceutical" OR
 TITLE_ABS:"biopharmaceuticals"
 OR TITLE_ABS:"biologics" OR TITLE_ABS:"monoclonal antibody" OR
 TITLE_ABS:"monoclonal antib
 odies" OR TITLE_ABS:"vaccine" OR TITLE_ABS:"vaccines" OR
 TITLE_ABS:"therapeutic protein" O
 R TITLE_ABS:"tableting" OR TITLE_ABS:"tablet press" OR
 TITLE_ABS:"tablet compression" OR T
 ITLE_ABS:"tablet coating" OR TITLE_ABS:"direct compression" OR
 TITLE_ABS:"roller compactio
 n" OR TITLE_ABS:"blend uniformity" OR TITLE_ABS:"content uniformity"
 OR TITLE_ABS:"process
 analytical technology") AND PUB_YEAR:[2018 TO 2026] NOT
 PUB_TYPE:"review" NOT SRC:PPR NOT
 SRC:ETH

OpenAlex (works, filter title_and_abstract.search)

Started 2026-09-23T13:58:04.390Z; finished 2026-09-23T13:59:08.337Z; count reported by the
 source 4738; records returned 4744.

("machine learning" OR "deep learning" OR "neural network" OR "neural
 networks" OR "random
 forest" OR "support vector machine" OR "support vector regression" OR
 "gradient boosting"
 OR "XGBoost" OR "artificial intelligence" OR "decision tree" OR "k-
 nearest neighbor" OR "
 gaussian process regression" OR "ensemble learning") AND
 ("pharmaceutical manufacturing" O
 R "drug manufacturing" OR "biopharmaceutical manufacturing" OR

"biomanufacturing" OR "bioprocess" OR "bioprocessing" OR "bioreactor" OR "process analytical technology" OR "continuous manufacturing" OR "pharmaceutical process" OR "pharmaceutical quality" OR "quality by design" OR "critical quality attribute" OR "critical quality attributes" OR "critical process parameter" OR "critical process parameters" OR "granulation" OR "tableting" OR "tablet press" OR "tablet compression" OR "tablet coating" OR "direct compression" OR "roller compaction" OR "blend uniformity" OR "content uniformity" OR "spray drying" OR "hot melt extrusion" OR "freeze drying" OR "lyophilization" OR "crystallization" OR "batch release" OR "real-time release" OR "release testing" OR "stability testing" OR "shelf life" OR "visual inspection" OR "quality control" OR "process monitoring" OR "fault detection" OR "dissolution") AND ("pharmaceutical" OR "pharmaceuticals" OR "drug product" OR "drug products" OR "drug substance" OR "active pharmaceutical ingredient" OR "biopharmaceutical" OR "biopharmaceuticals" OR "biologics" OR "monoclonal antibody" OR "monoclonal antibodies" OR "vaccine" OR "vaccines" OR "therapeutic protein" OR "tableting" OR "tablet press" OR "tablet compression" OR "tablet coating" OR "direct compression" OR "roller compaction" OR "blend uniformity" OR "content uniformity" OR "process analytical technology")

Semantic Scholar (Academic Graph API, paper search)

Started 2026-09-23T13:58:04.390Z; finished 2026-09-23T13:58:19.580Z; count reported by the source 3474; records returned 3474.

("machine learning" | "deep learning" | "neural network" | "neural networks" | "random forest" | "support vector machine" | "support vector regression" | "gradient boosting" | "XGBoost" | "artificial intelligence" | "decision tree" | "k-nearest neighbor" | "gaussian process regression" | "ensemble learning") + ("pharmaceutical manufacturing" | "drug manufacturing" | "biopharmaceutical manufacturing" | "biomanufacturing" | "bioprocess" | "bioprocessing" | "bioreactor" | "process analytical technology" | "continuous manufacturing" | "pharmaceutical process" | "pharmaceutical quality" | "quality by design" | "critical quality

attribute" | "critical quality attributes" | "critical process parameter" | "critical process parameters" | "granulation" | "tableting" | "tablet press" | "tablet compression" | "tablet coating" | "direct compression" | "roller compaction" | "blend uniformity" | "content uniformity" | "spray drying" | "hot melt extrusion" | "freeze drying" | "lyophilization" | "crystallization" | "batch release" | "real-time release" | "release testing" | "stability testing" | "shelf life" | "visual inspection" | "quality control" | "process monitoring" | "fault detection" | "dissolution") + ("pharmaceutical" | "pharmaceuticals" | "drug product" | "drug products" | "drug substance" | "active pharmaceutical ingredient" | "biopharmaceutical" | "biopharmaceuticals" | "biologics" | "monoclonal antibody" | "monoclonal antibodies" | "vaccine" | "vaccines" | "therapeutic protein" | "tableting" | "tablet press" | "tablet compression" | "tablet coating" | "direct compression" | "roller compaction" | "blend uniformity" | "content uniformity" | "process analytical technology")

DOAJ (article search, bibjson.abstract; split in four because of the URL length limit)

Started 2026-09-23T13:58:04.391Z; finished 2026-09-23T13:58:06.971Z; count reported by the source [155, 77, 53, 89]; records returned 374.

Query part 1:

```
bibjson.abstract:("machine learning" OR "deep learning" OR "neural network" OR "neural networks" OR "random forest" OR "support vector machine" OR "support vector regression" OR "gradient boosting" OR "XGBoost" OR "artificial intelligence" OR "decision tree" OR "k-nearest neighbor" OR "gaussian process regression" OR "ensemble learning") AND bibjson.abstract:("pharmaceutical manufacturing" OR "bioprocess" OR "continuous manufacturing" OR "critical quality attribute" OR "granulation" OR "tablet coating" OR "content uniformity" OR "lyophilization" OR "release testing" OR "quality control") AND bibjson.abstract:("pharmaceutical" OR "pharmaceuticals" OR "drug product" OR "drug products" OR "drug substance" OR "active pharmaceutical ingredient" OR "biopharmaceutical" OR
```

"biopharmaceuticals" OR "biologics"
" OR "monoclonal antibody" OR "monoclonal antibodies" OR "vaccine" OR
"vaccines" OR "therapeutic protein" OR "tableting" OR "tablet press" OR "tablet
compression" OR "tablet coating" OR "direct compression" OR "roller compaction" OR "blend
uniformity" OR "content uniformity" OR "process analytical technology") AND bibjson.year:[2018 TO
2026]

Query part 2:

bibjson.abstract:("machine learning" OR "deep learning" OR "neural
network" OR "neural networks" OR "random forest" OR "support vector machine" OR "support
vector regression" OR "gradient boosting" OR "XGBoost" OR "artificial intelligence" OR
"decision tree" OR "k-nearest neighbor" OR "gaussian process regression" OR "ensemble learning")
AND bibjson.abstract
:("drug manufacturing" OR "bioprocessing" OR "pharmaceutical process"
OR "critical quality attributes" OR "tableting" OR "direct compression" OR "spray drying"
OR "crystallization" OR "stability testing" OR "process monitoring") AND bibjson.abstract:
("pharmaceutical" OR
"pharmaceuticals" OR "drug product" OR "drug products" OR "drug
substance" OR "active pharmaceutical ingredient" OR "biopharmaceutical" OR "biopharmaceuticals"
OR "biologics" OR "monoclonal antibody" OR "monoclonal antibodies" OR "vaccine" OR
"vaccines" OR "therapeutic protein" OR "tableting" OR "tablet press" OR "tablet compression" OR
"tablet coating" OR
"direct compression" OR "roller compaction" OR "blend uniformity" OR
"content uniformity"
OR "process analytical technology") AND bibjson.year:[2018 TO 2026]

Query part 3:

bibjson.abstract:("machine learning" OR "deep learning" OR "neural
network" OR "neural networks" OR "random forest" OR "support vector machine" OR "support
vector regression" OR "gradient boosting" OR "XGBoost" OR "artificial intelligence" OR
"decision tree" OR "k-nearest neighbor" OR "gaussian process regression" OR "ensemble learning")
AND bibjson.abstract
:("biopharmaceutical manufacturing" OR "bioreactor" OR "pharmaceutical
quality" OR "critical process parameter" OR "tablet press" OR "roller compaction" OR "hot

```

melt extrusion" OR
"batch release" OR "shelf life" OR "fault detection") AND
bibjson.abstract:("pharmaceutica
l" OR "pharmaceuticals" OR "drug product" OR "drug products" OR "drug
substance" OR "activ
e pharmaceutical ingredient" OR "biopharmaceutical" OR
"biopharmaceuticals" OR "biologics"
OR "monoclonal antibody" OR "monoclonal antibodies" OR "vaccine" OR
"vaccines" OR "therap
eutic protein" OR "tableting" OR "tablet press" OR "tablet
compression" OR "tablet coating
" OR "direct compression" OR "roller compaction" OR "blend uniformity"
OR "content uniform
ity" OR "process analytical technology") AND bibjson.year:[2018 TO
2026]

```

Query part 4:

```

bibjson.abstract:("machine learning" OR "deep learning" OR "neural
network" OR "neural net
works" OR "random forest" OR "support vector machine" OR "support
vector regression" OR "g
radient boosting" OR "XGBoost" OR "artificial intelligence" OR
"decision tree" OR "k-neare
st neighbor" OR "gaussian process regression" OR "ensemble learning")
AND bibjson.abstract
:("biomanufacturing" OR "process analytical technology" OR "quality by
design" OR "critica
l process parameters" OR "tablet compression" OR "blend uniformity" OR
"freeze drying" OR
"real-time release" OR "visual inspection" OR "dissolution") AND
bibjson.abstract:("pharma
ceutical" OR "pharmaceuticals" OR "drug product" OR "drug products" OR
"drug substance" OR
"active pharmaceutical ingredient" OR "biopharmaceutical" OR
"biopharmaceuticals" OR "bio
logics" OR "monoclonal antibody" OR "monoclonal antibodies" OR
"vaccine" OR "vaccines" OR
"therapeutic protein" OR "tableting" OR "tablet press" OR "tablet
compression" OR "tablet
coating" OR "direct compression" OR "roller compaction" OR "blend
uniformity" OR "content
uniformity" OR "process analytical technology") AND bibjson.year:[2018
TO 2026]

```

Line breaks inside the query blocks are for display only; each query is one string.