



European Network
of Cancer Registries



European
Commission



Krebsregister
Baden-Württemberg

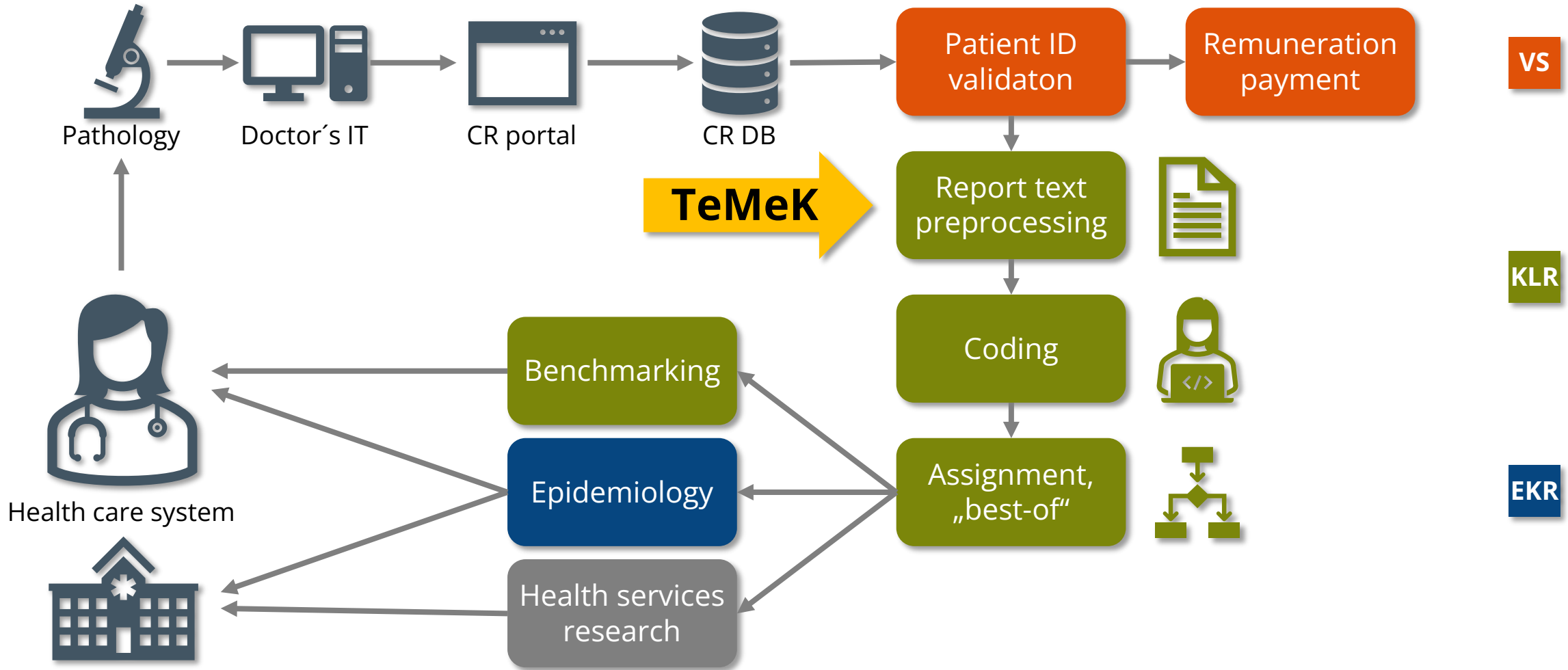
Textmining of pathology reports for uniform classifications (TeMeK)

ENCR-IACR 2023 Scientific Conference, Granada, Spain
Escuela Andaluza de Salud Pública (EASP) – Cancer Registry of Granada

2023-11-13, 13:30-17:00, Workshop 4:
IT Tools and novel AI approaches for cancer registration

Prof. Dr. med. **Marco Halber**, Dipl.-Betriebswirt (FH)
<https://temek.de>, <https://www.krebsregister-bw.de>

Processing pathology reports



Free text classification example

Diagnose nach Paraffineinbettung:
Die makroskopische und histologische Begutachtung ergibt:

Lobektomiepräparat (Oberlappen links) mit einem mäßig differenzierten, prädominant azinär wachsenden Adenokarzinom (G2) mit Haem- und Lymphangiosis carcinomatosa sowie einem max. 1,2 cm messenden Satellitenherd. Keine Infiltration der Pleura visceralis. Ein tumorfreier Lymphknoten (0/1; LK 12- 14). Das tumorferne Lungengewebe mit feinbullösem Lungenemphysem.

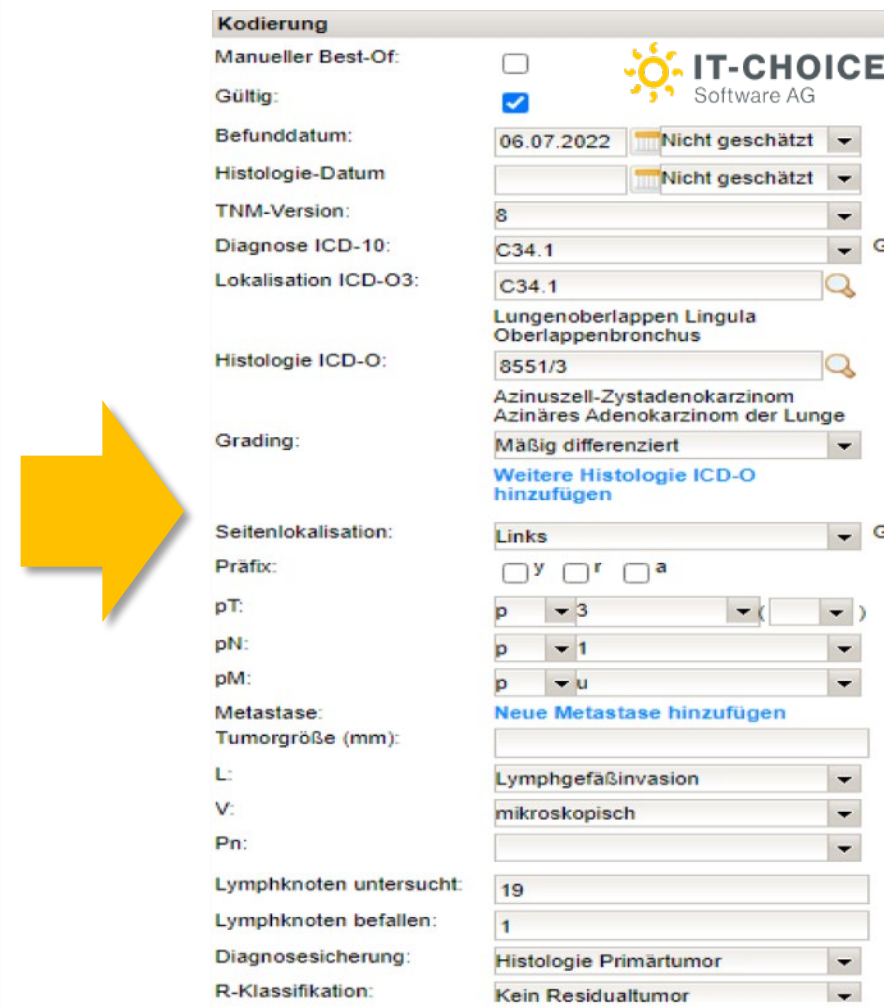
Unter Berücksichtigung der Paralleleinsendung R/2022/3278 ergibt sich folgende

Tumorklassifikation:

TNM (8. Aufl.): pT3, pN1 (1/19), L1, V1, PL0

Grading: G2

R-Klassifikation (lokal): R0



Kodierung

Manueller Best-Of:

Gültig: IT-CHOICE Software AG

Befunddatum: 06.07.2022 Nicht geschätzt

Histologie-Datum: Nicht geschätzt

TNM-Version: 8

Diagnose ICD-10: C34.1

Lokalisation ICD-O3: C34.1
Lungenoberlappen Lingula Oberlappenbronchus

Histologie ICD-O: 8551/3
Azinuszell-Zystadenokarzinom Azinäres Adenokarzinom der Lunge

Grading: Mäßig differenziert
Weitere Histologie ICD-O hinzufügen

Seitenlokalisierung: Links

Präfix: y r a

pT: p 3

pN: p 1

pM: p u

Metastase: Neue Metastase hinzufügen

Tumorgröße (mm):

L: Lymphgefäßinvasion

V: mikroskopisch

Pn:

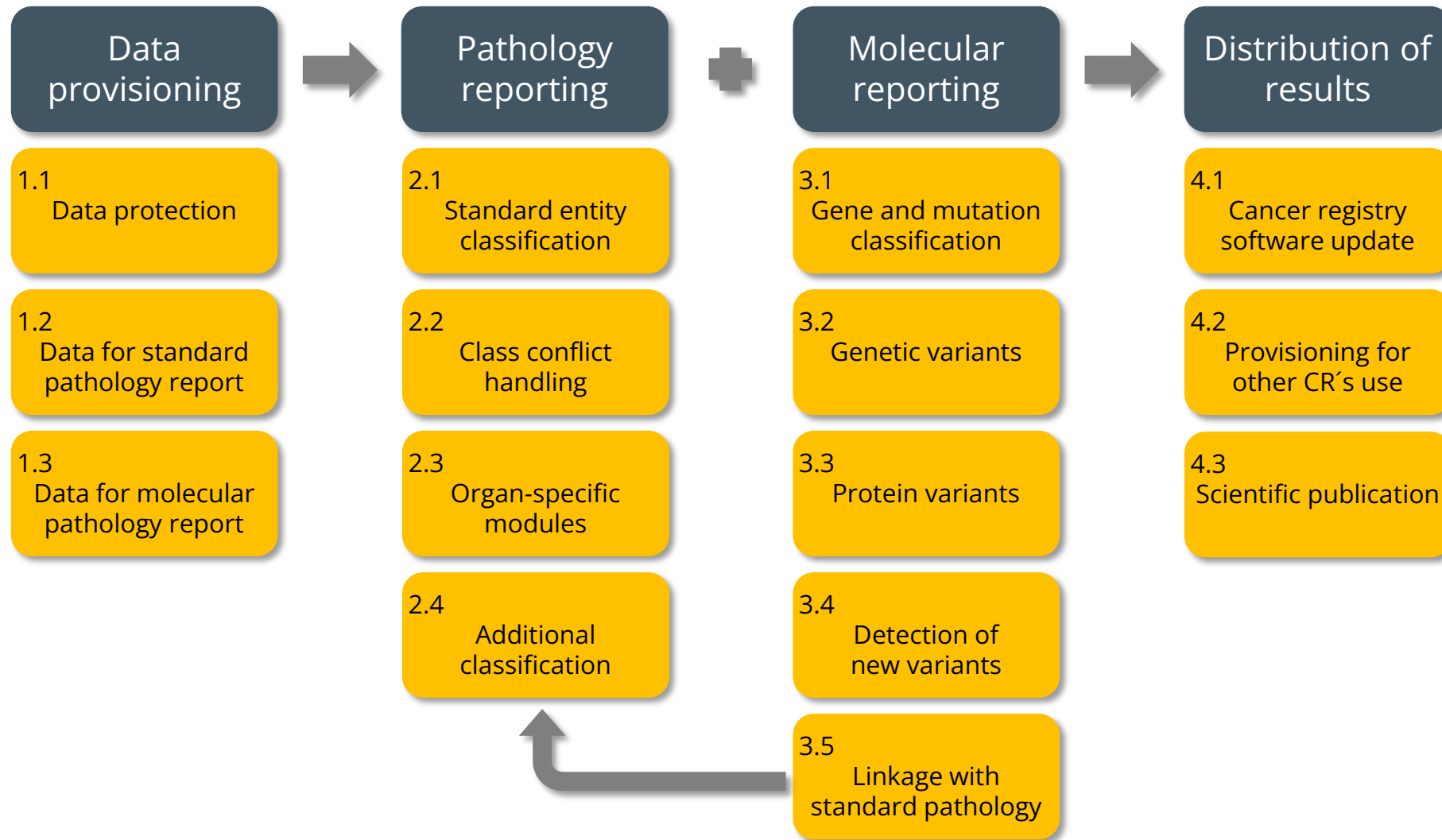
Lymphknoten untersucht: 19

Lymphknoten befallen: 1

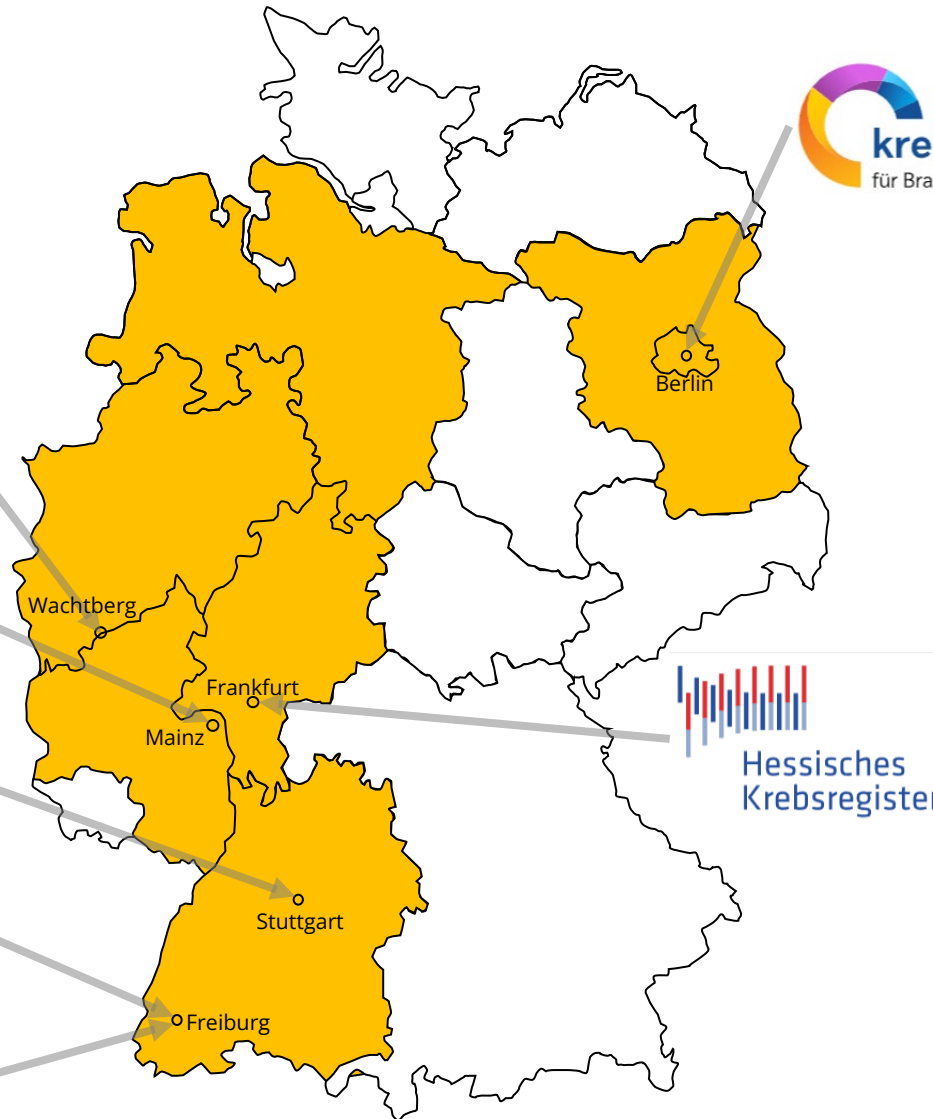
Diagnosesicherung: Histologie Primärtumor

R-Klassifikation: Kein Residualtumor

TeMeK work packages



TeMeK partners



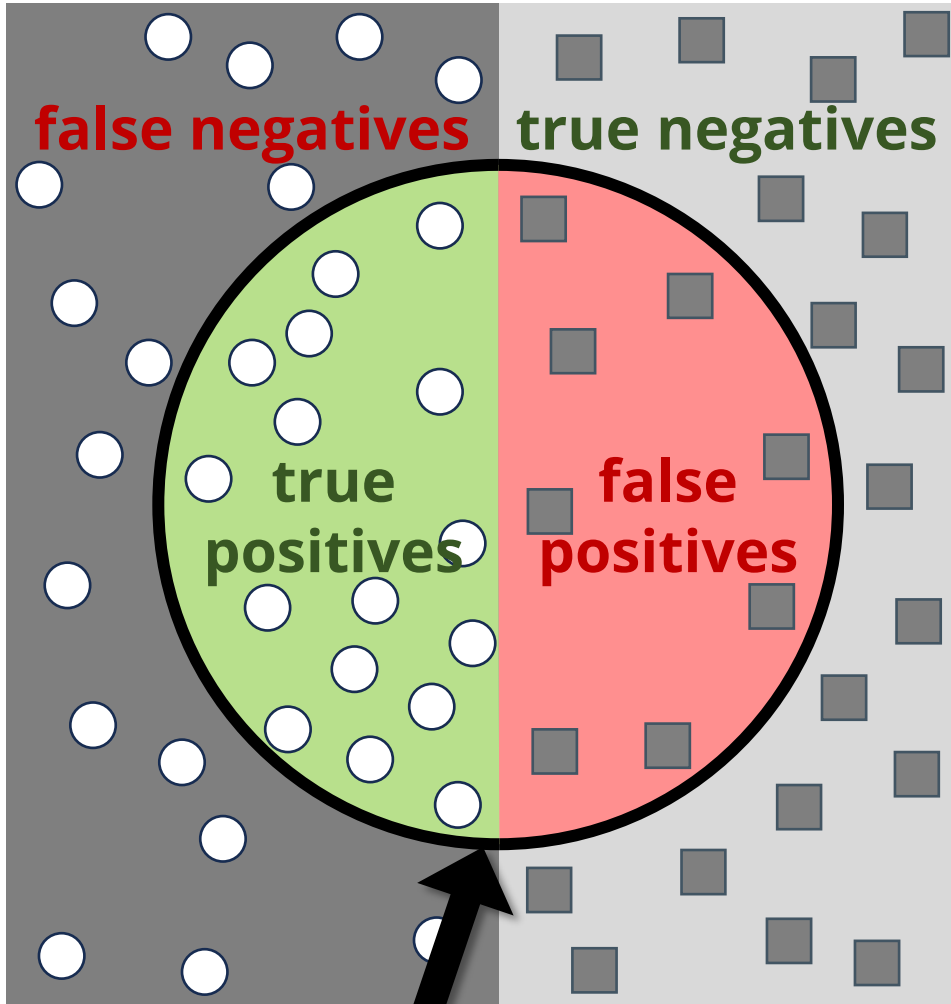
- **Krebsregister Baden-Württemberg**
- **Klinisches Krebsregister für Brandenburg und Berlin**
- **Hessisches Krebsregister**
- **Krebsregister Rheinland-Pfalz**
- **Tumorzentrum Freiburg – CCCF**
- **Averbis GmbH, Freiburg**
- **Fraunhofer FKIE**

Additional pathology reports kindly supplied by

- **Klinisches Krebsregister Niedersachsen**
- **Krebsregister Nordrhein-Westfalen**

What is success?

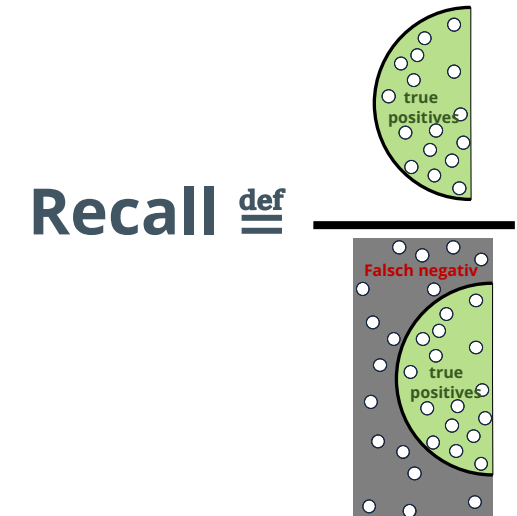
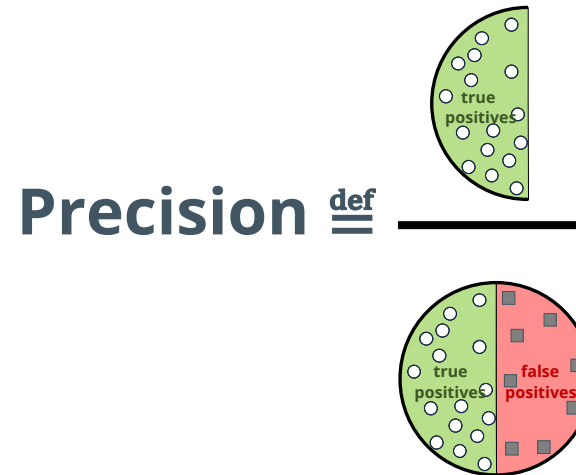
○ = relevant ■ = nicht relevant



Result = relevant („retrieved“) = positive

How many **retrieved** items are **relevant**?

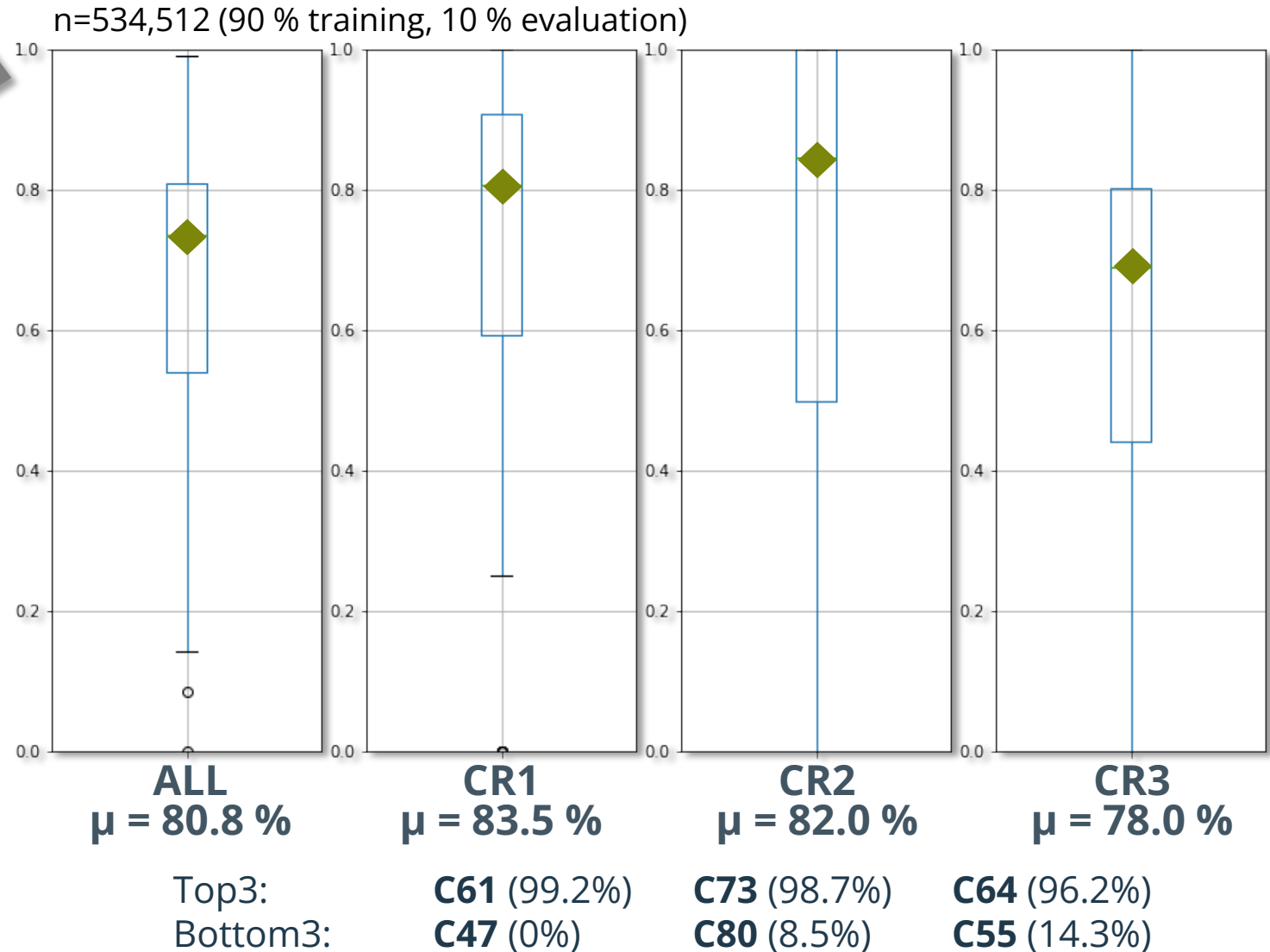
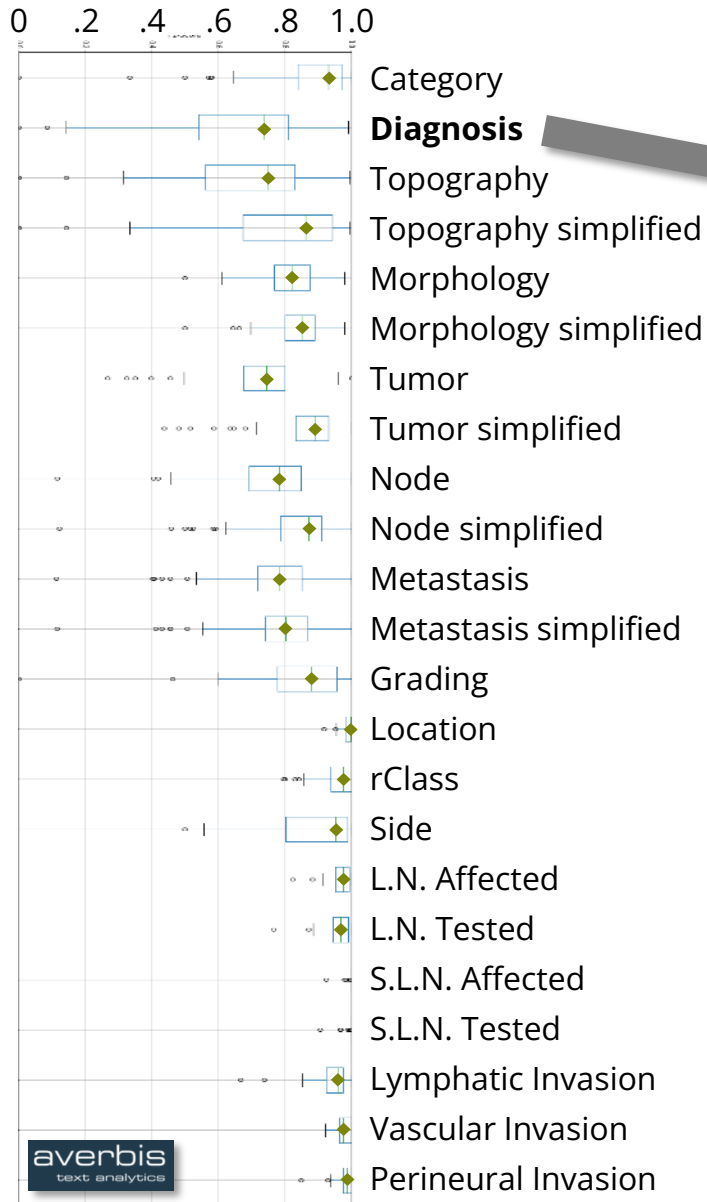
How many **relevant** items are **retrieved**?



F1 $\stackrel{\text{def}}{=} \text{harmonic mean of precision \& recall}$

$$\bar{x}_{\text{harm}} = \frac{n}{\frac{1}{x_1} + \dots + \frac{1}{x_n}}$$

Preliminary results: F1 scores



Das TeMeK-Projekt

Wer?

Warum?

Wie?

Perspektiven

TextMining von Meldungstexten für einheitliche Klassifikationen



Im Projekt „TeMeK“ untersuchen wir **Methoden der Künstlichen Intelligenz (KI)** für einheitliche Informationen aus komplexem Freitext für **klinische Krebsregister**.

„Erkenntnisse aus unentdecktem Wissen“

ChatGPT 2023-01-29

(Antwort auf „entwerfe einen slogan für ein textmining-projekt“)



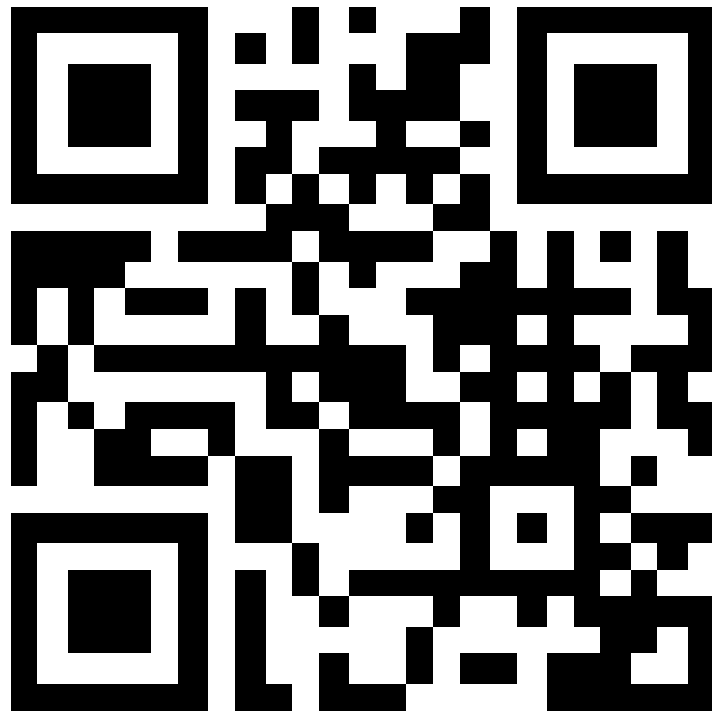
<https://temek.de>

Förderung: **Bundesministerium für Gesundheit** als [Ressortforschung](#) im Handlungsfeld [Digitalisierung](#),
Forschungsschwerpunkt „[Krebsregisterdaten zusammenführen und intelligent nutzen](#)“.

Impressum

Datenschutz

Thank You



<https://www.thelaend.de/en/>

