Improving the identification of heart failure in the EHR in general practice using HFpEF scoring systems and natural language processing

Ruben Styl, Universiteit Antwerpen

Evi Noppe, Universiteit Antwerpen

Promotor: prof. dr. Miek Smeets

Co-promotor: prof. dr. Andreas Gevaert

Master of Family Medicine

Masterproef Huisartsgeneeskunde
Academiejaar: 2025 - 2026

Deze masterproef is een examendocument dat niet werd gecorrigeerd voor eventueel vastgestelde fouten. Zonder voorafgaande schriftelijke toestemming van zowel de promotor(en) als de auteur(s) is overnemen, kopiëren, gebruiken of realiseren van deze uitgave of gedeelten ervan verboden. Voor aanvragen tot of informatie i.v.m. het overnemen en/of gebruik en/of realisatie van gedeelten uit deze publicatie, wendt u tot de universiteit waaraan de auteur is ingeschreven.

Voorafgaande schriftelijke toestemming van de promotor(en) is eveneens vereist voor het aanwenden van de in dit afstudeerwerk beschreven (originele) methoden, producten, schakelingen en programma's voor industrieel of commercieel nut en voor de inzending van deze publicatie ter deelname aan wetenschappelijke prijzen of wedstrijden.

Abstract

Background. Heart failure (HF) affects over 64 million people worldwide and is frequently underdiagnosed in primary care, particularly HF with preserved ejection fraction (HFpEF). The OSCAR-HF audit-and-feedback strategy improves identification but is restricted to structured data and is time-intensive. This study aimed to evaluate two complementary strategies: (i) a natural language processing (NLP) tool combining clinical concept extraction, normalisation to SNOMED-CT and LOINC, and direct application of the OSCAR diagnostic flowchart and H2FPEF score; and (ii) the addition of H2FPEF case-finding to the OSCAR flowchart for HFpEF-specific identification.

Methods. We conducted a monocentric study in two primary care practices, including 3109 patients aged ≥40 years across two electronic health record (EHR) systems. The OSCAR extended audit identified 358 patients at risk for HF (audit population). Two primary researchers manually applied the OSCAR flowchart and HFpEF case-finding to this group (reference standard). The NLP tool was applied to all 3109 patients without practice-specific training, returning a structured set of OSCAR and H2FPEF parameters together with extracted symptoms and clinical signs. Discordant cases between NLP and manual review, and NLP-positive cases outside the audit population, were reviewed by a multidisciplinary expert panel (general practitioner, cardiologist, geriatrician) according to the European Society of Cardiology (ESC) definition of HF.

Results. Within the audit population, NLP showed almost perfect agreement with manual review for OSCAR criteria ($\kappa = 0.812$) and substantial agreement for HFpEF case-finding ($\kappa = 0.704$). Of 15 OSCAR-discordant cases (NLP-positive, manual-negative), the expert panel confirmed 11 (73.3%) as HF, indicating that manual review missed a substantial proportion of true cases. For HFpEF case-finding, only 5 of 19 NLP-only cases (26.3%) were confirmed. This lower rate is consistent with the role of H2FPEF as a probabilistic case-finding score. Outside the audit population, NLP detected 36 additional patients with possible HF. Expert-panel confirmation was substantially higher when NLP also extracted clinical symptoms or signs from free-text documentation (6 of 14; 42.9%) than when none were extracted (1 of 22; 4.5%), illustrating the diagnostic value of accessing unstructured data that structured audits cannot reach. NLP processed all 3109 patients in approximately 23 hours, compared with 20.6 hours of manual review for the 358-patient audit population alone. In addition, within the audit population, adding HFPEF case-finding to the OSCAR flowchart identified 25 additional patients with probable HFpEF; the panel confirmed HFpEF in 7 (28.0%) and classified the remainder as high risk for HFpEF, warranting referral if symptoms develop.

Conclusion. NLP enabled comprehensive HF case-finding across the full patient record, identifying cases missed by both structured audits and manual review whilst having the potential to substantially reduce the manual workload that currently limits routine audit-and-feedback methods. Combining HFpEF-specific case-finding with the OSCAR flowchart improves identification of patients at high risk for HFpEF, although formal diagnosis still requires clinical evaluation. Together, these complementary approaches may help address the underdetection of HF, particularly HFpEF, in primary care. Following further validation in larger populations and evaluation of clinical impact, this approach could form the foundation of a learning health system for continuous monitoring of HF identification and quality of care.

Abstract **3**
Introduction **5**
Methods **7**
1. Design and setting 7
2. Procedures 7
2.1 Extended audit and feedback method 7
2.2 OSCAR flowchart objectivation and HFpEF case-finding method 7
2.3 NLP method 9
2.4 Expert panel validation 10
3. Data collection 12
4. Outcomes 13
4.1 Research question 1 - NLP method: 13
4.1.1 Primary outcome 13
4.1.2 Secondary outcome 13
4.1.3 Tertiary outcome 13
4.2 Research question 2 - HFpEF case-finding: 13
5. Statistical analysis 14
6. Ethics & General Data Protection Regulation (GDPR) 14
Results **15**
1. HF audit population characteristics and audit results 15
2. RQ 1: Diagnostic accuracy and added yield of NLP compared to manual method 16
2.1 Primary outcome: diagnostic accuracy of NLP compared to manual review 16
2.1.1 OSCAR flowchart 16
2.1.2 HFpEF case-finding method 17
2.2 Secondary outcome: expert panel validation of discordant cases within the HF audit population 18
2.3 Exploratory analysis: NLP findings outside the HF audit population 19
2.4 Tertiary outcome: time efficiency 20
3. RQ 2: Manual HFpEF case-finding using the H2FPEF score 21
Discussion **22**
Strengths and limitations 26
Conclusion **27**
Citations **28**
Attachments **31**
1. Basic audit: queries used to search for registered HF patients in the EHR 31
2 Extended audit: queries used to search for additional patients within the EHR 31
3. Oscar flowchart and HFpEF case-finding method details 32
4. H2FPEF-score 32
5. Symptoms and clinical signs 33

Introduction

Heart failure (HF) affects over 64 million people worldwide and represents a growing public health concern.(1) General practitioners (GPs) play a central role in managing these predominantly older, multimorbid patients.(2,3) Accurate identification is essential for patient management and developing learning health systems.(3,4) However, identifying HF in primary care remains challenging. Diagnostic tools such as echocardiography and biomarkers are often underused,(2,5)and diagnostic coding is inaccurate.(2,6) The Belgian OSCAR-HF study demonstrated that an extended audit identified 74% more HF patients than were initially recorded. In that process, nearly half of pre-existing coded HF diagnoses were not confirmed, and only 67% of newly identified HF cases by GP's could be objectified using diagnostic criteria.(2,7)

The extended audit and feedback strategy applied in the OSCAR-HF study proved effective but also exhibited important limitations.(2) The approach relied exclusively on structured data within general practice records, thereby excluding clinically relevant information contained in cardiology reports. Moreover, key clinical features, such as symptoms and signs, are often sparsely or inconsistently documented in primary care and could not be reliably extracted. As a result, confirmation of HF diagnoses depended on GPs clinical judgment, introducing a risk of misclassification. Identification of heart failure with preserved ejection fraction (HFpEF) is particularly challenging in this context. Symptoms frequently manifest only during exertion, while resting examinations may appear normal (5), and objective diagnostic criteria are often missing or inadequately documented.(2) Notably, in the OSCAR-HF study, HF diagnoses were absent from specialist correspondence in up to 66% of HFpEF cases.(2) Although a diagnostic flowchart based on expert panel consensus was developed to standardize HF diagnosis, it was not specifically tailored to address the unique diagnostic complexities of HFpEF. On top of this, the strategy demanded extensive manual review and time investment(2,7), limiting routine implementation.

Two approaches offer potential solutions. Large Language Models (LLMs) can extract clinical information from unstructured electronic health record (EHR) data (8) and normalize clinical concepts to standardized terminologies like systematized nomenclature of medicine-clinical terms (SNOMED-CT).(9–11) Additionally, the H2FPEF score enables HFpEF case-finding using clinical variables combined with echocardiographic parameters.(12) Despite their promise, these approaches have yet to be validated in primary care. Most natural language processing (NLP) studies for HF detection were hospital-based, where documentation patterns differ substantially from general practice (13,14) and the H2FPEF score was primarily evaluated in hospital settings.(15,16) Combining these approaches for automated HF identification in primary care has not yet been studied.

We addressed these gaps by implementing two complementary approaches in Flemish primary care. First, we evaluated HFpEF-specific case-finding based on the H2FPEF score. Second, we developed and tested an NLP tool combining automated clinical concept extraction and normalisation with both the OSCAR diagnostic flowchart and HFpEF case-finding. We compared these approaches to traditional manual audit methods in patients aged ≥40 years across two general practices.

Methods

1. Design and setting

We conducted this monocentric study for a duration of 6 months in two primary care practices in Flanders, Belgium using two EHR systems 'CareConnect' (Corilus NV, Belgium) and 'Daktari' (CMG, Belgium).

All patients aged 40 years or older with an active patient file were included. Patients who were identified by the OSCAR extended audit method constituted the 'HF audit' population. As described below, additional patients were detected with probable HF but not identified by the OSCAR extended audit method; these constituted the 'undetected HF' population.

2. Procedures

2.1 Extended audit and feedback method

We performed an extended audit and feedback method as performed and validated in the OSCAR studies.(2,7)

First, we performed a basic and extended audit in the statistics module of the EHRs, generating a list of possible HF patients. The basic audit queried registered HF diagnoses (coded or free text searches) (see attachment 1). The extended audit applied additional predefined coded and free text search strings reflecting established HF risk factors and typical HF medications.(2) The search strategy was further optimized using the 10 search terms identified as essential through machine learning analysis (7), to which we added two additional typical HF medications (see attachment 2).

Next, the audit list of possible HF patients was reviewed by the participating GPs and classified as HF, possible HF or non-HF. The GPs first received a presentation on HF by the primary researchers (PRs), including the definition of HF and diagnostic criteria such as signs, symptoms and echocardiographic parameters. For each individual study patient, T0 was defined as the moment of evaluation by the participating GPs. For patients outside the HF audit population, T0 was defined as the date when the audit was performed (24-06-25 for practice 1 and 25-08-25 for practice 2). For all the following procedures, only data available in the EHR at the time of evaluation (T0) was used.

2.2 OSCAR flowchart objectivation and HFpEF case-finding method

The PRs manually applied the validated OSCAR diagnostic flowchart (Figure 1) to the HF audit population to identify patients with objective criteria for an HF diagnosis. In contrast to the original OSCAR-HF study design, in which manual objectivation of the diagnostic flowchart is performed only on patients flagged as HF by GP feedback, we applied the manual flowchart to the entire HF audit population. This allowed for a direct method comparison between manual and NLP review across the same population, independent of clinical feedback by the GP. The flowchart was specifically developed and validated for use in primary care, and patients meeting its criteria were considered to have definite

HF by an expert panel in prior OSCAR-HF studies.(2) It applies established diagnostic criteria across six sequential steps, incorporating echocardiographic findings and biomarker data, to enable systematic objectivation of HF diagnoses within the primary care EHR. During the study, the expert panel refined several criteria by applying stricter thresholds. Detailed definitions of the criteria and the adjustments are described in attachment 3.

The PRs manually reviewed each patient file (consultation notes, specialist letters, laboratory results, medication lists and coded problem lists) for criteria corresponding to the six sequential steps. The time required to complete each patient evaluation was measured.

Validating an HFpEF diagnosis is particularly challenging in primary care. The OSCAR flowchart requires high NT-proBNP values or objective echocardiographic evidence of diastolic dysfunction. Unfortunately these parameters are frequently absent or incompletely documented.(2,5) As symptoms often manifest primarily during exertion, a standard resting echocardiogram is insufficient for establishing the diagnosis.(2) A low objectivation rate using the OSCAR flowchart therefore does not necessarily reflect diagnostic error, but rather the absence of documented objective evidence at the time of evaluation. This limitation is particularly relevant in older primary care populations, where HFpEF is the predominant HF phenotype.(2, 17)

In order to improve identification of HFpEF cases, we used a case-finding method based on the established H2FPEF score.(12) This score was chosen because it estimates HFpEF probability using clinical variables and echocardiographic parameters, without requiring invasive haemodynamic measurement. The score incorporates six variables: obesity (Body Mass Index (BMI) >30 kg/m², 2 points), hypertension (use of two or more antihypertensive medications, 1 point), atrial fibrillation (3 points), pulmonary hypertension (estimated pulmonary artery systolic pressure (PASP) >35 mmHg, 1 point), age >60 years (1 point), and elevated left ventricular filling pressure (echocardiographic E/e' ratio >9, 1 point). A score of ≥6 points indicates a probability of HFpEF exceeding 90% in patients with unexplained dyspnea.(12) This method (Figure 1) was applied to all patients in the HF audit population during manual chart review, simultaneously with the OSCAR flowchart. Patients meeting a H2FPEF score of ≥6, but not identified by the OSCAR flowchart, were considered probable HFpEF cases and referred for expert panel evaluation.

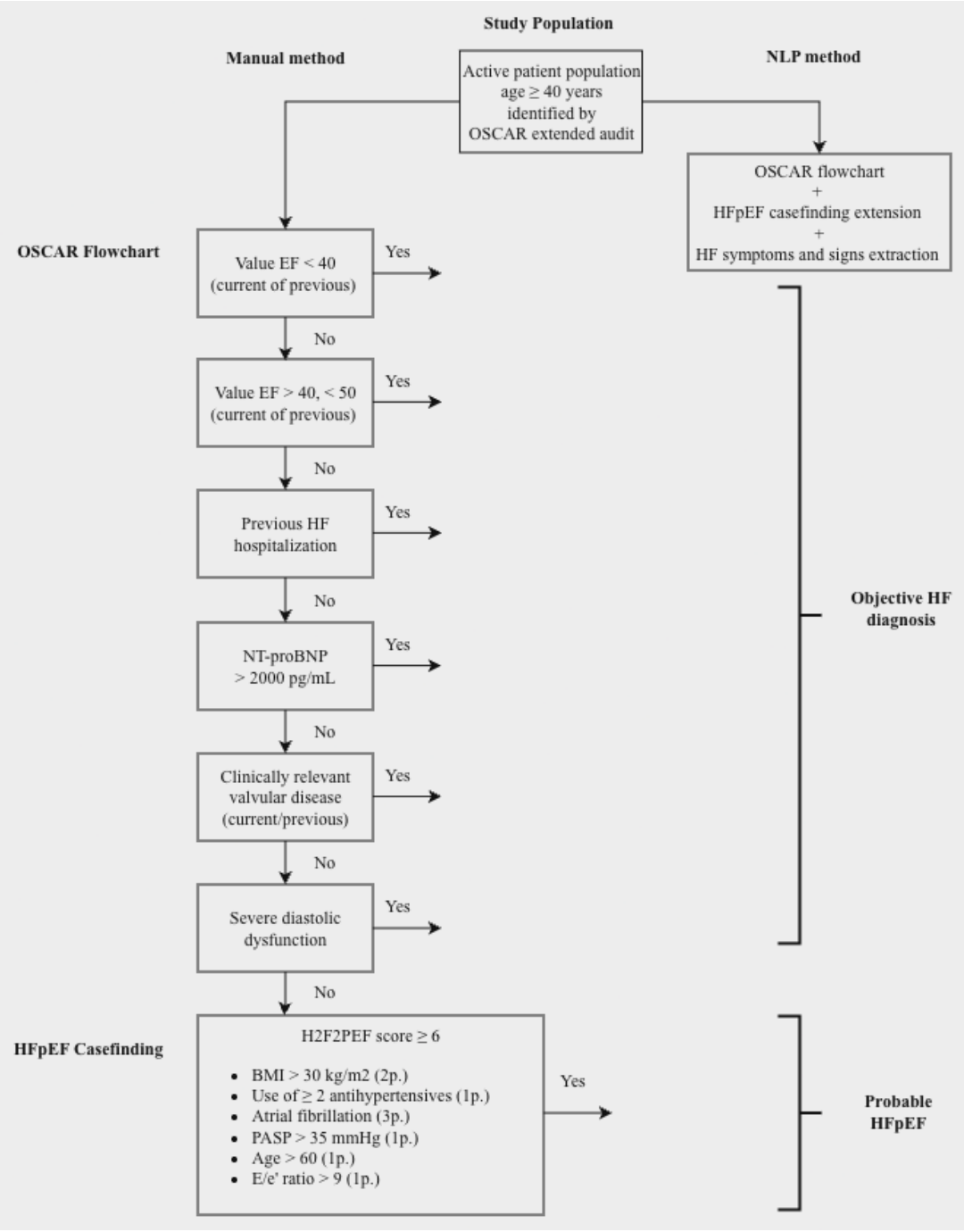


Figure 1: *Oscar flowchart and HFpEF case-finding method*

2.3 NLP method

Active files of patients aged 40 and older were extracted from the EHR and pseudonymised by removing all personal identifiers (see Section 3 'Data collection'). Entries in the EHR dated after T0 were removed from each patient file. The GPs uploaded the pseudonymized data sets to a dedicated platform (Co-Medic, Belgium, co-developed by one of the PRs) for processing using a NLP pipeline. An existing NLP model performed clinical concept recognition and concept normalization to SNOMED CT and Logical Observation Identifiers Names and Codes (LOINC) without retraining or fine-tuning on data from the participating practices. The normalisation process mapped various phrasings of the same

clinical concept, such as "EF 38%", "ejection fraction of 38 percent", and "LVEF 38%", to a single coded entity. Negation and clinical (un)certainty were also handled as part of the normalisation process. Both manual and NLP methods could access the same EHR data (consultation notes, specialist letters, lab results, medication list, and coded problem list). The platform returned a structured set of parameters corresponding to the OSCAR-flowchart and the HFpEF case-finding method for each patient, including backlinks to the original data sources. Additionally, the platform returned a list of symptoms and signs relevant to HF (attachment 5) which were used during the expert panel validation (Section 2.4).

2.4 Expert panel validation

Expert panel validation was performed on three patient groups (Figure 2).

First, within the HF audit population, all patients identified as probable HFpEF by manual case-finding but not meeting OSCAR criteria (H2FPEF+/OSCAR-) were reviewed to determine what proportion could be confirmed as HFpEF.

Second, also within the HF audit population, discordant cases defined as patients classified as OSCAR-positive or probable HFpEF by NLP but not confirmed by manual review (NLP+/Manual-) were reviewed to distinguish false negatives of manual review from false positives of the NLP method, and to estimate the added diagnostic yield of NLP.

Third, a selection of patients outside the HF audit population identified as OSCAR-positive or probable HFpEF by NLP were reviewed to estimate the proportion of patients missed by the audit but with objective HF.

The expert panel comprised a GP with expertise in HF, a cardiologist, and a geriatrician. For each patient the panelists reviewed the patient record alongside a list of clinical signs and symptoms relevant to HF which were extracted by the NLP tool. The panel recorded a diagnostic conclusion for each case based on the European Society of Cardiology (ESC) definition of HF (18) and, where applicable, the reason for rejection.

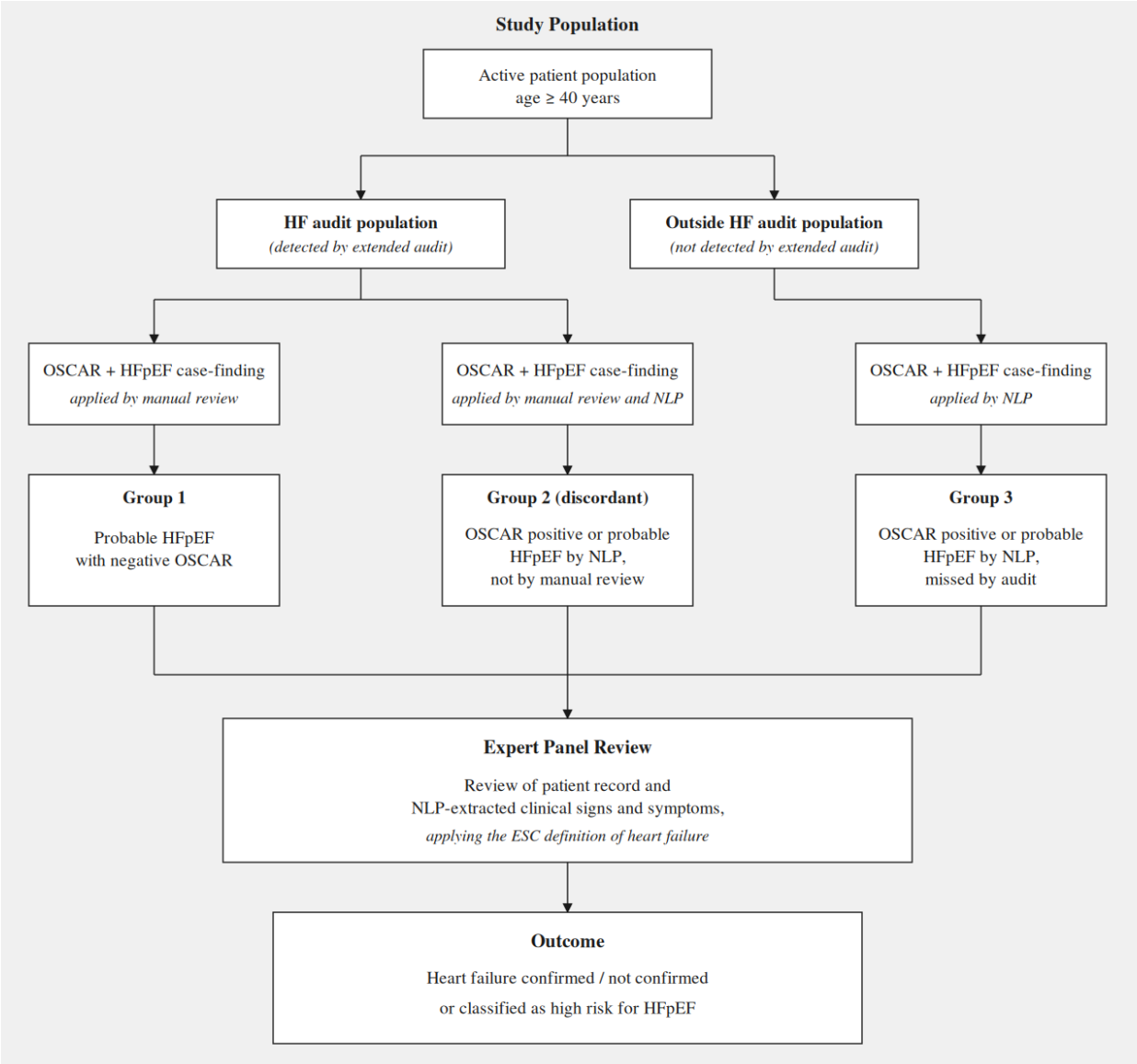


Figure 2: *Groups reviewed by the expert panel*

3. Data collection

For the NLP method, patient records were extracted from the EHR. We pseudonymised all records before transfer and analysis, using a locally operated pseudonymisation tool. Each patient received a practice-specific study identifier (Practice 1: PT1000X; Practice 2: PT2000X).

The pseudonymisation measures which were taken:

- replacing patient name and national identification number with the study identifier;
- converting date of birth to age in years;
- masking other personal and contact data (e.g., address/telephone) and masking data of healthcare professionals mentioned in the records with dummy values;

A separate mapping table which linked identifiers (name, date of birth) to the study identifier was created and stored separately under restricted access by the study promotor. The pseudonymised study database was stored on University of Antwerp infrastructure with restricted access for authorised study personnel. Pseudonymised files uploaded to the Co-Medic platform were stored for the duration of the analysis and deleted at the end of the study period.

4. Outcomes

4.1 Research question 1 - NLP method:

The primary objective of research question (RQ) 1 is to evaluate how the implementation of an NLP tool impacts diagnostic accuracy and efficiency in identifying HF patients compared to traditional manual audit methods.

4.1.1 Primary outcome

The diagnostic accuracy of the NLP method was measured by sensitivity, specificity, positive and negative predictive values, and the Cohen's kappa coefficient for inter-method agreement. The manual method was used as the reference standard for both the OSCAR flowchart and the HFpEF case-finding method.

4.1.2 Secondary outcome

The added diagnostic yield of NLP compared to manual detection was evaluated by the expert panel. Discordant cases (patients identified as positive by NLP but not confirmed by manual review) were reviewed by the panel:

1. Patients identified as OSCAR-positive by NLP but not by manual chart review.
2. Patients classified as probable HFpEF by NLP but not by manual chart review.

Outside the HF audit population (exploratory analysis)

Among all patients aged ≥40 years not identified by the extended audit, NLP was applied to identify patients meeting OSCAR criteria and/or having probable HFpEF. All NLP-positive patients outside the HF audit population were referred for expert panel validation to estimate the proportion of audit-missed patients with confirmable HF.

4.1.3 Tertiary outcome

The time efficiency was assessed descriptively by the time required for

1. Manual application of the OSCAR flowchart and HFpEF case-finding method by the PRs
2. Review of the audit population by the GPs (HF, possible HF or non-HF).

4.2 Research question 2 - HFpEF case-finding:

To determine whether a HFpEF-specific diagnostic strategy can improve the identification of HFpEF cases compared to the OSCAR diagnostic flowchart alone, we evaluated the following outcomes:

1. The absolute number of additional cases with probable HFpEF found by the manual case-finding method compared to the OSCAR-flowchart alone;
2. The percentage of additional cases regarded to have HFpEF by the expert panel.

5. Statistical analysis

Continuous variables were reported as mean ± standard deviation (SD) when the data were normally distributed. We used median with interquartile range (IQR) when the data were non-normally distributed. The distribution of data was assessed by the Shapiro-Wilk test. Categorical variables are reported as absolute frequencies and percentages.

The diagnostic accuracy of the NLP method relative to manual review was assessed using sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), each with 95% confidence intervals (CI) calculated using the Wilson score method. The manual method served as the reference standard for all primary diagnostic accuracy analyses. Inter-method agreement was quantified using Cohen's kappa coefficient (κ), interpreted accordingly: <0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial and >0.80 almost perfect agreement. Cohen's kappa coefficients were reported with 95% confidence intervals.

A two-sided p-value of <0.05 was considered statistically significant. All analyses were conducted in SPSS version 31.

6. Ethics & General Data Protection Regulation (GDPR)

Ethics committee approval was obtained from the University Hospitals Antwerp Ethics Committee in March 2025 (Project ID 7192).

Results

1. HF audit population characteristics and audit results

The characteristics of participating practices and patients are shown in Table 1. The GPs cared for a total patient population of 6825 patients. The proportion of patients aged 40 years and older was 3109 patients (45.6%). The search for a registered HF diagnosis in GPs' EHRs identified 44 patients, representing a registered HF prevalence of 1.4% among patients aged ≥40 years. The extended audit identified an additional 314 patients not previously identified, resulting in a total HF audit population of 358 patients (11.5% of the ≥40 population).

One patient meeting the age criterion (≥40 years) was excluded because the EHR file could not be exported (from practice 2). One patient identified by the extended audit was excluded due to a faulty match related to documentation of a psychiatric decompensation episode (from practice 2).

Table 1. *HF audit population characteristics and audit results*

Characteristic	
Participating practices	2
Total registered patients	6,825
Patients ≥40 years, n (%)	3109 (45.6%)
HF audit population (% of ≥40 years)	**358 (11.5%)**
Coded HF diagnoses	44 (1.4%)
Identified by extended audit only	314 (10.1%)
Age (years), mean ± SD	72.4 ± 12.5
Male, n (%)	205 (57.3%)

2. RQ 1: Diagnostic accuracy and added yield of NLP compared to manual method

2.1 Primary outcome: diagnostic accuracy of NLP compared to manual review

2.1.1 OSCAR flowchart

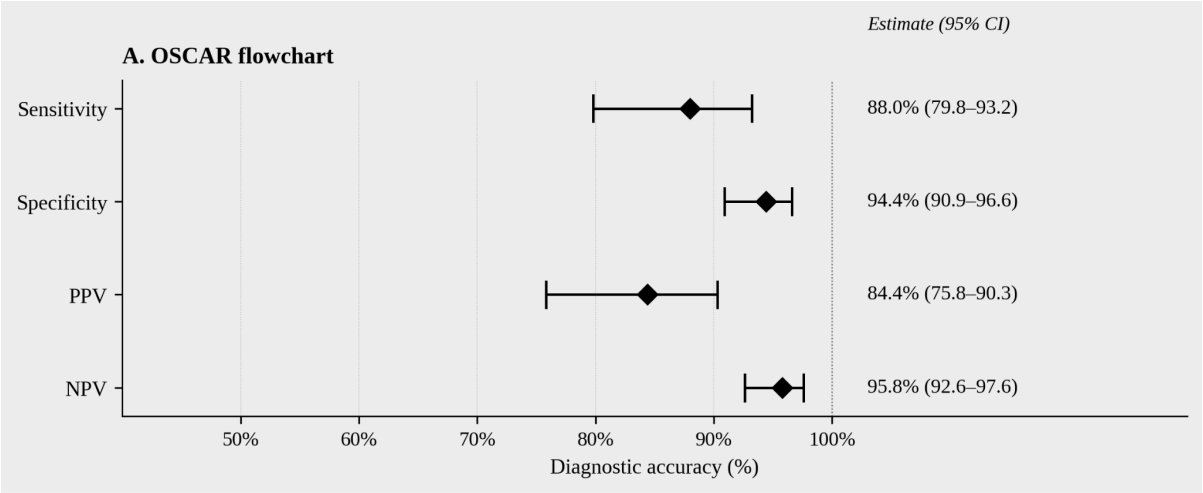
Using manual review as the reference standard, NLP identified 96 patients (26.8%) meeting OSCAR criteria, compared to 92 (25.7%) by manual review. NLP detected 15 additional cases (4.2%) not identified during manual review, while manual review identified 11 cases (3.1%) not detected by NLP (Table 2a). Agreement between methods was almost perfect (Cohen's κ = 0.812, 95% CI 0.74-0.88; p < .001) (Figure 4).

NLP demonstrated high sensitivity of 88.0% (95% CI 79.8-93.2%) and specificity of 94.4% (95% CI 90.9-96.6%), with a positive predictive value (PPV) of 84.4% (95% CI 75.8-90.3%) and negative predictive value (NPV) of 95.8% (92.6-97.6%), compared to the manual review (Figure 3a).

Table 2a. *OSCAR criteria: NLP versus manual review (n = 358)*

	Manual +	Manual -	Total
NLP +	81	15	96
NLP -	11	251	262
Total	**92**	**266**	**358**

Figure 3a. *Diagnostic accuracy of NLP for OSCAR criteria, compared to manual review (n = 358)*



Squares represent point estimates and horizontal lines represent 95% confidence intervals for sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV).

2.1.2 HFpEF case-finding method

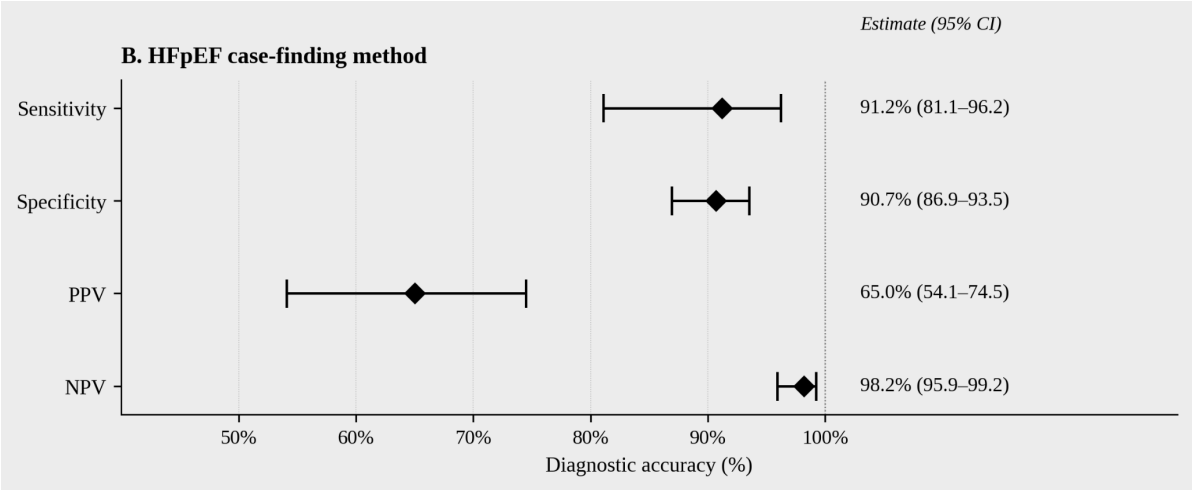
For HFpEF case-finding, NLP identified 80 patients (22.3%) with probable HFpEF compared to 57 (15.9%) by manual review, with good agreement (Cohen's κ = 0.704, 95% CI 0.61-0.80; p < .001) (Figure 4). NLP detected 28 additional probable cases not identified manually, while manual review identified 5 cases not detected by NLP (Table 2b).

NLP demonstrated sensitivity of 91.2% (95% CI: 81.1-96.2%) and specificity of 90.7% (95% CI: 86.9-93.5%), with a PPV of 65.0% (95% CI: 54.1-74.5%) and NPV of 98.2% (95% CI: 95.9-99.2%) (Figure 3b).

Table 2b. *HFpEF case-finding: NLP versus manual review (n = 358)*

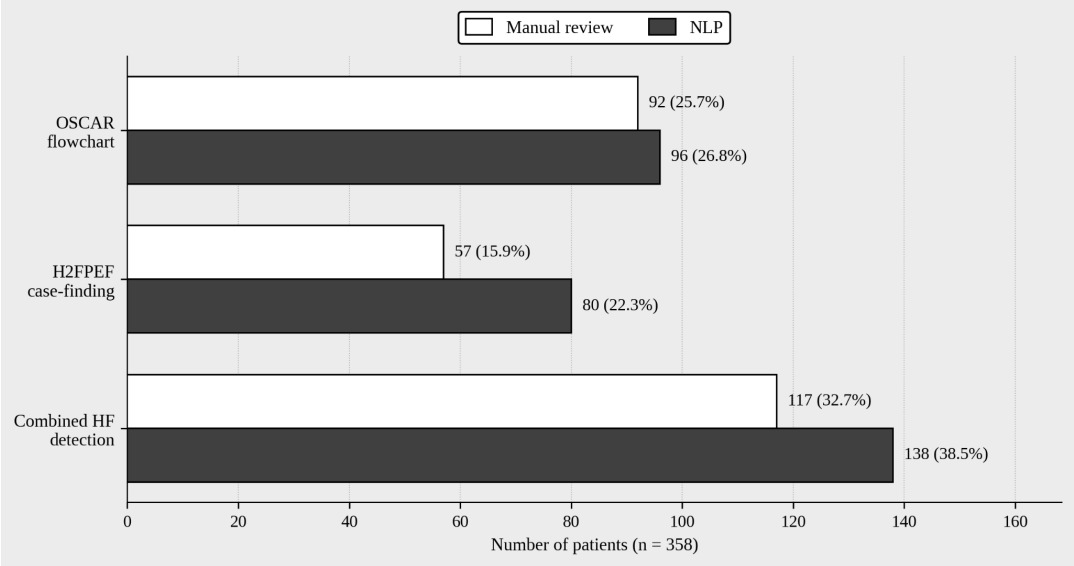
	Manual +	Manual -	Total
NLP +	52	28	80
NLP -	5	273	278
Total	**57**	**301**	**358**

Figure 3b. *Diagnostic accuracy of NLP for HFpEF case-finding, compared to manual review (n = 358)*



Squares represent point estimates and horizontal lines represent 95% confidence intervals for sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV).

Figure 4. *HF case detection by manual review and NLP within the HF audit population (n = 358).*



Bars: patient counts per method (% of audit population). Combined HF = OSCAR+ and/or probable HFpEF

2.2 Secondary outcome: expert panel validation of discordant cases within the HF audit population

Of the 15 patients identified as OSCAR-positive by NLP but not by manual review, expert panel validation confirmed 11 (73.3%) as having HF. Of the 19 patients classified as probable HFpEF by NLP but not by manual review, expert panel validation confirmed 5 (26.3%) as having HF (Table 3).

Table 3. *Expert panel validation of discordant cases (NLP+/Manual-)*

Discordant group	n	Confirmed HF n (%)	Not confirmed n (%)
OSCAR	15	11 (73.3%)	4 (26.7%)
H2FPEF	19	5 (26.3%)	14 (73.7%)

2.3 Exploratory analysis: NLP findings outside the HF audit population

Among all patients aged ≥40 years outside the HF audit population (n=2751) NLP identified 32 patients (1.2%) with objective OSCAR criteria, representing 25.0% of all NLP-detected OSCAR-positive cases. NLP identified 6 patients (0.2%) with probable HFpEF, of whom 4 (0.1%) were not identified by OSCAR criteria and thus represent additional cases detected exclusively through HFpEF case-finding. In total, 36 patients (1.3%) with any HF criteria (OSCAR or HFpEF) were identified outside the HF audit population (Table 4).

Table 4. *NLP detection across patient population ≥40 years*

NLP finding	HF audit pop. (n = 358)	Outside audit pop. (n = 2751)	Total (n = 3109)	Captured by audit
OSCAR+	96	32	128 (4.1%)	75.0%
H2FPEF+	80	6	86 (2.8%)	93.0%
of which OSCAR-	42	4	46 (1.5%)	91.3%
Combined HF	**138**	**36**	**174 (5.6%)**	**79.3%**

Expert panel validation of the 36 NLP-detected cases outside the HF audit population confirmed 7 (19.4%) as having HF. Confirmation rates differed strongly by symptom extraction: 6 of 14 (42.9%) patients in whom NLP identified HF symptoms or signs were confirmed by the expert panel, compared to only 1 of 22 (4.5%) patients without NLP-extracted symptoms (Table 5).

Table 5. *Expert panel validation of NLP findings outside HF audit population (n = 36)*

NLP finding	n	Confirmed HF n (%)	Not confirmed n (%)
With signs or symptoms	14	6 (42.9%)	8 (57.1%)
Without signs or symptoms	22	1 (4.5%)	21 (95.5%)
Total	**36**	7 (19.4%)	29 (80.6%)

2.4 Tertiary outcome: time efficiency

The manual application of the OSCAR flowchart and HFpEF case-finding method required a mean of 3.5 ± 2.1 minutes per patient (median 3.0 minutes; range 1-20 minutes), totaling approximately 20.6 hours for the HF audit population of 358 patients (Table 6).

GP clinical assessment of the audit list required a mean of 0.8 ± 0.6 minutes per patient (median 0.7 minutes; range 0.02 - 4.6 minutes), which totaled approximately 4.7 hours for 358 patients (Table 6).

For comparison, the NLP pipeline processed all 3,109 patients in approximately 23 hours, corresponding to a mean of approximately 27 seconds per patient. Applied to the audit population, manual review and GP assessment combined required approximately 25 hours, while NLP processing of the equivalent number of patients would correspond to less than 3 hours of computation time.

Table 6. *Time efficiency (n = 358)*

Metric	Manual review	GP clinical assessment
Mean time per patient (min)	3.5 ± 2.1	0.8
Median time per patient (min)	3.0	0.7
Range (min)	1-20	0.02 - 4.6
Total time	**~ 1239 min (20.6 hours)**	**~ 282 min (4.7 hours)**

3. RQ 2: Manual HFpEF case-finding using the H2FPEF score

Of the 358 patients in the HF audit population, the manual application of the OSCAR flowchart identified 92 patients with objective evidence of HF. The H2FPEF score identified 57 patients with probable HFpEF (H2FPEF score ≥ 6). Of these 57 probable cases, 32 already met OSCAR criteria. 25 patients had a high H2FPEF score but did not meet OSCAR criteria. Validation by an expert panel confirmed HFpEF in 7 of these 25 additional cases (Table 7).

Table 7. *HFpEF case detection: OSCAR vs HFpEF case-finding (n = 358)*

Detection method	n	% of HF audit pop. (n = 358)	Confirmed by expert panel n (%)
OSCAR+	92	25.7%	-
H2FPEF+	57	15.9%	-
of which OSCAR-	25	7.0%	7 (28.0%)
Combined detection	117	32.7%	-

Discussion

The main findings of this study are twofold. First, in 3109 patients across two primary care practices, an NLP tool performing automated concept extraction and normalisation established strong agreement with the manual reference standard within a preselected population at risk for HF identified by a structured audit, while additionally identifying cases not detected by manual review. Beyond this preselected population, NLP further identified patients with probable HF who would otherwise have remained undetected. Second, in the preselected population, a HFpEF-specific case-finding method using the H2FPEF score was able to identify a group at high risk for HFpEF. However, formal diagnosis still required clinical evaluation using echocardiographic parameters in the presence of symptoms and/or clinical signs.

First, we evaluated how the implementation of an NLP tool impacts diagnostic accuracy and efficiency in identifying HF patients compared to traditional audit methods. NLP achieved almost perfect agreement with manual review for OSCAR criteria ($\kappa = 0.812$) and substantial agreement for HFpEF case-finding ($\kappa = 0.704$), with high sensitivity and specificity for both OSCAR criteria (88.0%, 94.4%) and HFpEF case-finding (91.2%, 90.7%). Positive predictive values were 84.4% and 65.0% respectively. The lower PPV for HFpEF case-finding is expected, as H2FPEF is a case-finding score rather than a diagnostic criterion (further discussed below). For OSCAR, 11 of 15 discordant cases were confirmed as HF by the expert panel, meaning most 'false positives' were HF cases missed by manual review. This indicates that both calculated PPV's underestimate the true diagnostic performance of NLP, and that manual review itself has limited sensitivity as a reference standard. Outside the preselected population, NLP detected an additional 36 patients with possible HF, of whom 7 (19.4%) were confirmed.

The OSCAR-HF studies have shown that an extended audit and feedback intervention can considerably improve HF identification in primary care, decreasing both under- and overregistration.(2,7) However, feedback by GPs and/or manual validation of patients identified by the audit against objective criteria are time-intensive and remain practical obstacles to implementation in routine practice.(7) Our findings provide a first evaluation of NLP as a possible solution to this bottleneck in Belgian primary care.

Our findings about the performance of our NLP method are consistent with prior studies demonstrating LLM-based approaches for extracting HF-related concepts from unstructured EHR data. Zero-shot extraction of HF symptoms and signs from hospital notes achieved 90.6% PPV and 100% sensitivity using GPT-4, outperforming traditional rule-based and machine learning methods.(8) Existing NLP studies for HF detection have been conducted mainly in hospital settings.(13,14) Our study extends this evidence to a primary care context with clinical documentation in the Dutch language.

A second distinctive element of our NLP method is the normalisation of extracted clinical concepts to SNOMED-CT and LOINC codes, which enables their downstream application in diagnostic algorithms. Negation and clinical (un)certainty are also handled as part of the normalisation process. This allows the logic of the OSCAR flowchart and the H2FPEF score to be applied systematically. Previous studies have shown that the incorporation of SNOMED-CT improves NLP performance (19) and the normalisation of concepts enabled prediction models to generalise across different Dutch primary care EHR systems.(20) To our knowledge, the combination of LLM-based concept extraction and

terminology normalisation to perform a validated HF case-finding algorithm has not previously been described in primary care.

Outside the preselected population, NLP identified 36 patients meeting OSCAR criteria or probable HFpEF, of whom 7 (19.4%) were confirmed by the expert panel. Confirmation rates differed strongly by symptom presence: among patients with documented HF symptoms or signs, 6 of 14 (42.9%) were confirmed, compared to only 1 of 22 (4.5%) without documented symptoms. This underscores the value of an NLP approach that captures symptoms and clinical signs from unstructured documentation in addition to structured diagnostic criteria, given the central role of symptoms in the formal definition of HF.(18) That NLP detected confirmable HF cases outside the audit population reflects a limitation of current audit capabilities in Belgian primary care EHRs. The extended audit is restricted to coded data and free-text entries within the structured care element fields and the medication list, and cannot access clinical concepts embedded in specialist letters, imaging reports, or consultation notes integrated into the patient record. Yet much of the most crucial information resides within these unstructured documents. Two Dutch studies have illustrated this gap: in one study, only 13% of concepts extracted from primary care free text had a corresponding structured code,(21) and another study showed that text-based HF retrieval across almost one million records remained confined to episode description fields, unable to access broader clinical information.(6) Our NLP tool allowed for the processing of the complete patient record, including consultation notes and specialist correspondence, accessing clinical information that cannot be processed using structured audit methods.

However, NLP did not detect all manually identified cases: 11 OSCAR-positive and 5 probable HFpEF patients were missed. We did not formally evaluate whether these false negatives are due to errors in manual review or in NLP concept extraction and normalisation.

GP classification of the audit list (358 patients) took approximately 4.7 hours. For this group, the application of the OSCAR flowchart and HFpEF case-finding by the PRs required 20.6 hours of chart review. The processing of all 3109 patients by the NLP pipeline took approximately 23 hours.

Altogether, these findings point towards a possible evolution in HF identification methods in primary care. Searching for registered diagnoses is practical but inaccurate, characterised by both over- and underregistration.(7) The extended audit and feedback intervention improves this but remains time-intensive as it is dependent on GP judgement and confined to structured data fields. NLP addresses both of these limitations by processing the full patient record and may provide the foundation for a learning health system in which HF identification and care quality are continuously monitored and improved in routine primary care.(22)

In the second part of this study, we examined whether the addition of a HFpEF-specific case-finding method to the OSCAR diagnostic flowchart can improve the identification of HFpEF cases. Of the 358 patients in the audit population, the OSCAR flowchart identified 92 patients (25.7%) with objective HF criteria. The H2FPEF score identified 25 additional patients (7.0%) with probable HFpEF (H2FPEF-score ≥6) who did not meet OSCAR criteria. Expert panel validation confirmed HFpEF in 7 of these 25 additional cases (28.0%). Following the ESC guidelines, HFpEF diagnosis was only confirmed in patients who had echocardiographic evidence and symptoms and clinical signs suggestive of HF. The 18 remaining patients classified as ‘no’, ‘possible' or ‘inconclusive’, were labelled by the panel as ‘high risk

for HFpEF', warranting referral for further diagnostic work-up in the presence of symptoms or clinical signs.

The H2FPEF score was developed by Reddy et al. (12) to assess HFpEF probability in patients with unexplained dyspnea, where a score ≥6 correlated to a probability of more than 90%. Its capability to differentiate patients with HFpEF from those without was later confirmed in hospital-based populations.(15,16) Our study extends the score to a primary care HF audit population in which dyspnea was not required as an inclusion criterion. As far as we know this is the first assessment of the H2FPEF score as a case-finding tool within a primary care HF audit. This shift in population has direct consequences for how the score should be interpreted. In Reddy's validation cohort, patients were already symptomatic and referred for third-line cardiology evaluation, so the pretest probability of HFpEF was high before the score was applied. In our study, we demonstrate that the H2FPEF score is also useful in a typical primary care population.

Following the ESC guidelines,(18) the expert panel confirmed 7 additional HFpEF cases that would have been missed by the OSCAR diagnostic flowchart. They only confirmed diagnosis in patients whose scores combined echocardiographic parameters (elevated E/e' ratio or estimated PASP > 35 mmHg) with documented symptoms or clinical signs. In patients who reached the ≥6 threshold on clinical variables alone (e.g. AF, age and BMI), HFpEF diagnosis could not be confirmed. Echocardiographic parameters are important as they show direct haemodynamic evidence of elevated filling pressures, while clinical risk factors indicate susceptibility to HFpEF rather than its presence. Both the HFA-PEFF diagnostic algorithm (23) and current ESC guidelines (18) prioritise objective haemodynamic and morphological evidence above clinical risk factors in diagnosing HFpEF.

For patients with a high H2FPEF score and symptoms or clinical signs, but without available echocardiographic parameters, the expert panel classified them as highly probable HFpEF without formal confirmation. This reflects a well-known challenge in primary care, where HFpEF predominantly affects very elderly patients in whom referral for further diagnostic investigation is often not pursued.(4) In such cases, probabilistic classification based on clinical variables, symptoms such as dyspnea, and age- and sex-adjusted NT-proBNP may be the most feasible approach, in line with the HFpEF-ABA score, which estimates HFpEF probability on clinical variables alone.(24)

Beyond diagnosis, elevated HFpEF probabilities may also have therapeutic value by suggesting earlier imaging or referral, and by identifying patients who may be at an earlier stage of disease. HFpEF typically develops after prolonged exposure to risk factors such as obesity, hypertension and systemic inflammation, resulting in progressive myocardial and extra-myocardial dysfunction. Patients identified at this stage may represent a target for preventive strategies, including SGLT2 inhibitors or weight reduction, although this approach still requires clinical validation.(25–27) Whether treatment decisions can be guided by probabilistic classification alone remains uncertain.

When echocardiography results were available, key parameters for diagnosing HFpEF required for the H2FPEF score, such as the E/e' ratio or estimated PASP, were often missing. This suggests either incomplete reporting measures relevant for HFpEF, or suboptimal acquisition during echocardiographic assessment. This is compatible with findings from the OSCAR-HF baseline study, in which a HF diagnosis was absent from specialist reports in 66% of HFpEF cases (2). This finding may

indicate a persistent underdiagnosis of HFpEF among cardiologists, or a structural gap in communication between cardiology and primary care.

As HF is a clinical syndrome, the expert panel could only formally confirm diagnosis in patients with documented symptoms or clinical signs suggestive of HF. This was sometimes difficult as many symptoms and signs are not specific for HF, occurring in other diseases such as COPD or pneumonia. In addition, symptoms found by the NLP model were not always linked with the timing of the echocardiography. Patients without symptoms or clinical signs suggestive for HF were classified as 'high risk for HFpEF'. In everyday practice, we suggest that these patients are flagged as such in the EHR so that they can be directly referred to a cardiologist or specialised dyspnea center in the occurrence of suggestive symptoms or signs.

Altogether, HFpEF remains a difficult diagnosis. A HFpEF-specific case-finding strategy in addition to the OSCAR flowchart can meaningfully enhance detection of patients at risk for HFpEF in a primary care population. Most of the audit population nevertheless remained without objective HF confirmation, which is not unexpected. As HF is a clinical syndrome, symptomatic patients reaching a H2FPEF score ≥3 with clinical risk factors alone, warrant referral for further diagnostic workup.(12) In symptomatic patients with a H2FPEF score of ≥6 in whom further investigation is not feasible, the high pre-test probability may justify a working diagnosis of HFpEF.(24) The clinical parameters needed for this assessment, including reported symptoms and echocardiographic parameters, are often missing or are largely documented in unstructured text within specialist letters and consultation notes and are therefore unavailable to structured audit methods alone. As shown above, we demonstrated that an NLP tool may overcome this obstacle by extracting and normalizing these parameters directly from free-text EHR data.

Strengths and limitations

Several strengths of this study were already mentioned above. The main strengths are the use of validated case-finding methods, the application of an NLP pipeline on data from two different EHR systems without practice-specific training and the validation of additional cases by a multidisciplinary expert panel. A predefined protocol with clear criteria was used for each flowchart step and H2FPEF parameter. Furthermore, the NLP pipeline developed for this study integrates LLM-based concept extraction with normalisation to SNOMED-CT and LOINC, which allows the output to be directly implemented within validated diagnostic and or case-finding algorithms. As mentioned above, the zero-shot performance that was realized across two different EHR systems assists the transferability of the method across different primary care data.

There are a few limitations, however. First, we conducted this study in two general practices with an HF audit population of 358 patients. This relatively small group reduces statistical precision and generalisability. Validation in a larger primary care population is required before solid conclusions can be drawn. Second, the efficiency of both diagnostic and case-finding methods depended on the completeness of the GP's EHR, which in turn depends on the information provided in cardiology reports and the quality of data registered by the GP, as earlier discussed. When relevant clinical information was never recorded in the EHR, HF cases remained undetectable, regardless of the case-finding method. Third, the patients identified by the H2FPEF score were not validated with invasive haemodynamic assessment, which remains the gold standard for confirming HFpEF. Fourth, the reference standard for the NLP comparison was manual review, but the results demonstrate that it is itself inexact, as NLP identified cases that were missed by manual review. Fifth, not all manually identified cases were detected by NLP. We did not formally evaluate whether these false negatives are due to errors in manual review or in NLP concept extraction and normalisation. Atypical abbreviations or specific phrasings in Flemish primary care notes are likely contributing factors, and could be addressed through model fine-tuning, although this would come at the cost of the zero-shot generalisability that characterises the current approach. In addition, the absence of SNOMED-CT or LOINC codes for some clinical concepts (such as elevated filling pressures) likely contributed to false negatives, and could be addressed through ontology expansion. Another limitation is that the clinical signs and symptoms supplied to the expert panel were extracted solely by the NLP and not independently confirmed by manual review. The extracted symptoms and signs were often non-specific for HF and could equally reflect other pathologies, such as dyspnoea attributable to pulmonary disease or peripheral oedema due to venous insufficiency. In addition, some negation errors were observed, in which negated mentions were extracted as positive findings. However, the expert panel had access to the source documents and reviewed them critically. Lastly, symptoms, signs, and diagnostic investigations were not temporally aligned, meaning that findings recorded some time apart could be combined as if part of a single clinical picture. These limitations are addressable through model fine-tuning, improved negation handling, and the incorporation of timestamp information. Sixth, because of the retrospective design we could not assess the clinical impact of the identification of additional HF cases, such as effects on treatment and patient outcomes. Seventh, reproducibility to other settings may be affected by differences in EHR structure, the organisation of primary care or communication between cardiology and general practice. Finally, our comparison between manual and NLP review was performed across the entire HF audit population, rather than on the GP-confirmed HF subset as in the OSCAR-HF studies. This approach was chosen to allow a direct comparison of NLP

performance against manual review on the same population. NLP may partially fulfil the role of GP feedback by extracting symptoms and clinical signs from free-text documentation, which the structured audit cannot reach. However, this study does not formally compare NLP-extracted symptoms against GP clinical judgement. As discussed earlier, our findings underscore that clinical evaluation remains essential, particularly for HFpEF.

Conclusion

In this primary care population, NLP showed strong agreement with manual chart review and additionally identified HF cases both within and beyond the audit population. Performing a HFpEF-specific case-finding strategy based on the H2FPEF score allowed for the identification of high risk HFpEF cases beyond those captured by the OSCAR flowchart alone.

Together, these findings indicate that a HFpEF-specific case-finding strategy based on the H2FPEF score can help address the persistent underdetection of HFpEF in primary care. Simultaneously, NLP may reduce the manual workload that currently limits the routine use of audit and feedback methods.

Further validation in larger primary care populations is needed. Prospective studies should assess the impact of this approach on treatment decisions and patient outcomes. If validated, this approach could pave the way for wider implementation in primary practice, where it could form the foundation of a learning health system for continuous monitoring of HF identification and quality of care.

Citations

1. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. Eur J Heart Fail. 2020;22(8):1342–56. DOI: 10.1002/ejhf.1858

2. Smeets M, Vaes B, Aertgeerts B, Raat W, Penders J, Vercammen J, et al. Impact of an extended audit on identifying heart failure patients in general practice: baseline results of the OSCAR-HF pilot study. ESC Heart Fail. 2020;7(6):3950–61. DOI: 10.1002/ehf2.12990

3. Smeets M, Henrard S, Aertgeerts B, Cools F, Janssens S, Vaes B. Methods to identify heart failure patients in general practice and their impact on patient characteristics: A systematic review. Int J Cardiol. 2018;257:199–206. DOI: 10.1016/j.ijcard.2017.06.108

4. Smeets M, Baldewijns K, Vaes B, Vandenhoudt H. Integration of Chronic Care in a Fragmented Healthcare System. Int J Health Policy Manag. 2022;12:7143. DOI: 10.34172/ijhpm.2022.7143

5. Gevaert AB, Kataria R, Zannad F, Sauer AJ, Damman K, Sharma K, et al. Heart failure with preserved ejection fraction: recent concepts in diagnosis, mechanisms and management. Heart Br Card Soc. 2022;108(17):1342–50. DOI: 10.1136/heartjnl-2021-319605

6. De Clercq L, Himmelreich JCL, Harskamp RE. Quality of heart failure registration in primary care: observations from 1 million electronic health records in the Amsterdam Metropolitan Area. Diagn Berl Ger. 2024; DOI: 10.1515/dx-2024-0009

7. Raat W, Smeets M, Henrard S, Aertgeerts B, Penders J, Droogne W, et al. Machine learning optimization of an electronic health record audit for heart failure in primary care. ESC Heart Fail. 2022;9(1):39–47. DOI: 10.1002/ehf2.13724

8. Workman T, Ahmed A, Sheriff H, Raman V, Zhang S, Shao Y, et al. ChatGPT-4 extraction of heart failure symptoms and signs from electronic health records. PubMed [Internet]. [cited 2025 Apr 15]; Available from: https://pubmed.ncbi.nlm.nih.gov/39442600/

9. Chang E, Mostafa J. The use of SNOMED CT, 2013-2020: a literature review. J Am Med Inform Assoc JAMIA. 2021;28(9):2017–26. DOI: 10.1093/jamia/ocab084

10. Seinen TM, Kors JA, van Mulligen EM, Fridgeirsson E, Rijnbeek PR. The added value of text from Dutch general practitioner notes in predictive modeling. J Am Med Inform Assoc JAMIA. 2023;30(12):1973–84. DOI: 10.1093/jamia/ocad160

11. Meystre S, Heider P, Cates A, Bastian G, Pittman T, Gentilin S, et al. Piloting an automated clinical trial eligibility surveillance and provider alert system based on artificial intelligence and standard data models. PubMed [Internet]. 2023 [cited 2025 Apr 15]; Available from: https://pubmed.ncbi.nlm.nih.gov/37041475/

12. Reddy YNV, Carter RE, Obokata M, Redfield MM, Borlaug BA. A Simple, Evidence-Based Approach to Help Guide Diagnosis of Heart Failure With Preserved Ejection Fraction. Circulation. 2018;138(9):861–70. DOI: 10.1161/CIRCULATIONAHA.118.034646

13. Reading Turchioe M, Volodarskiy A, Pathak J, Wright D, Tcheng J, Slotwiner D. Systematic review of current natural language processing methods and applications in

cardiology. PubMed [Internet]. 2022 [cited 2025 Apr 15]; Available from: https://pubmed.ncbi.nlm.nih.gov/34711662/

14. Levinson RT, Malinowski JR, Bielinski SJ, Rasmussen LV, Wilson MP, Manemann SM, et al. Identifying Heart Failure from Electronic Health Records: A Systematic Evidence Review. Cold Spring Harbor Laboratory Press; 2022; DOI: 10.1101/2021.02.01.21250933

15. Sun Y, Wang N, Li X, Zhang Y, Yang J, Tse G, et al. Predictive value of H2 FPEF score in patients with heart failure with preserved ejection fraction. PubMed [Internet]. [cited 2025 May 12]; Available from: https://pubmed.ncbi.nlm.nih.gov/33403825/

16. Reddy YNV, Kaye DM, Handoko ML, van de Bovenkamp AA, Tedford RJ, Keck C, et al. Diagnosis of Heart Failure With Preserved Ejection Fraction Among Patients With Unexplained Dyspnea. JAMA Cardiol. 2022;7(9):891–9. DOI: 10.1001/jamacardio.2022.1916

17. van Riet EES, Hoes AW, Limburg A, Landman MAJ, van der Hoeven H, Rutten FH. Prevalence of unrecognized heart failure in older persons with shortness of breath on exertion. Eur J Heart Fail. 2014;16(7):772–7. DOI: 10.1002/ejhf.110

18. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J. 2021;42(36):3599–726. DOI: 10.1093/eurheartj/ehab368

19. Chang E, Sung S. Use of SNOMED CT in Large Language Models: Scoping Review. PubMed [Internet]. 2024 [cited 2025 Apr 15]; Available from: https://pubmed.ncbi.nlm.nih.gov/39374057/

20. Seinen TM, Kors JA, van Mulligen EM, Fridgeirsson E, Rijnbeek PR. The added value of text from Dutch general practitioner notes in predictive modeling. J Am Med Inform Assoc. 2023;30(12):1973–84. DOI: 10.1093/jamia/ocad160

21. Seinen TM, Kors JA, van Mulligen EM, Rijnbeek PR. Using Structured Codes and Free-Text Notes to Measure Information Complementarity in Electronic Health Records: Feasibility and Validation Study. J Med Internet Res. 2025;27:e66910. DOI: 10.2196/66910

22. Nash D, Bhimani Z, Rayner J, Zwarenstein M. Learning health systems in primary care: a systematic scoping review. PubMed [Internet]. 2021 [cited 2026 Apr 14]; Available from: https://pubmed.ncbi.nlm.nih.gov/34162336/

23. Pieske B, Tschöpe C, de Boer RA, Fraser AG, Anker SD, Donal E, et al. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). Eur Heart J. 2019;40(40):3297–317. DOI: 10.1093/eurheartj/ehz641

24. Reddy YNV, Carter RE, Sundaram V, Kaye DM, Handoko ML, Tedford RJ, et al. An evidence-based screening tool for heart failure with preserved ejection fraction: the HFpEF-ABA score. Nat Med. Nature Publishing Group; 2024;30(8):2258–64. DOI: 10.1038/s41591-024-03140-1

25. Anker S, Butler J, Filippatos G, Ferreira J, Bocchi E, M B, et al. Empagliflozin in Heart Failure with a Preserved Ejection Fraction. PubMed [Internet]. 2021 [cited 2026 Apr 28]; Available from: https://pubmed.ncbi.nlm.nih.gov/34449189/

26. Solomon S, McMurray J, Claggett B, de Boer R, DeMets D, Af H, et al. Dapagliflozin in Heart Failure with Mildly Reduced or Preserved Ejection Fraction. PubMed [Internet]. 2022 [cited 2026 Apr 28]; Available from: https://pubmed.ncbi.nlm.nih.gov/36027570/

27. Borlaug B, Kitzman D, Davies M, Rasmussen S, E B, J B, et al. Semaglutide in HFpEF across obesity class and by body weight reduction: a prespecified analysis of the STEP-HFpEF trial. PubMed [Internet]. [cited 2026 Apr 28]; Available from: https://pubmed.ncbi.nlm.nih.gov/37635157/

Attachments

1. Basic audit: queries used to search for registered HF patients in the EHR

Heart Failure coded	ICPC-2-code: K77
Heart Failure free text	“hartfalen” OR “corfalen” OR “hartdecompensatie” OR “hartdecompensatie” OR “cordecompensatie” OR “NYHA” OR “LVfalen” OR “linkerventrikelfalen” OR “LVdysfunctie” OR “linkerventrikeldysfunctie” OR “LVdecompensatie” OR “linkerventrikeldecompensatie” OR “gedaalde EF” OR “gedaalde ejectiefractie” OR “gedaalde LVEF” OR “verminderde EF” OR “verminderde ejectiefractie” OR “verminderde LVEF” “decompensatie” NOT “psychische decompensatie”

2 Extended audit: queries used to search for additional patients within the EHR

1. Cardiomyopathy (free text)
2. Ischaemic heart disease (free text)
3. Mineralocorticoid receptor antagonists
4. Atrial fibrillation (coded)
5. Atrial fibrillation (free text)
6. Diuretics and angiotensin receptor blocker
7. Cardiomyopathy (coded)
8. Digoxin
9. Diuretics and ACE-inhibitor
10. Diuretics and beta blocker
11. Angiotensin receptor neprilysin inhibitor
12. Ivabradine
13. Empagliflozine
14. Dapagliflozine

3. Oscar flowchart and HFpEF case-finding method details

Clinically relevant valvular disease was defined as:

- Mitral valve stenosis, any grade
- Severe aortic valve stenosis, or aortic valve area < 1 cm^2
- Moderate-to-severe or severe mitral valve regurgitation (grade 2–3 or higher)*
- Moderate-to-severe or severe aortic valve regurgitation (grade 2–3 or higher)*

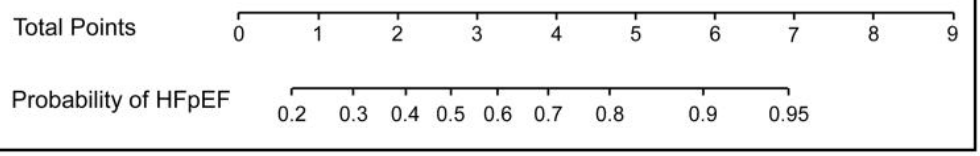
Severe diastolic dysfunction was defined as:

- Grade III or IV diastolic dysfunction*, and/or
- Restrictive filling pattern, and/or
- Elevated cardiac filling pressures, and/or
- E/e' > 14

* Threshold revised after expert panel review. Original criteria included moderate mitral and aortic valve regurgitation, and grade II diastolic dysfunction.

4. H2FPEF-score

	Clinical Variable	Values	Points
H_2	**H**eavy	Body mass index > 30 kg/m^2	2
	Hypertensive	2 or more antihypertensive medicines	1
F	Atrial **F**ibrillation	Paroxysmal or Persistent	3
P	**P**ulmonary Hypertension	Doppler Echocardiographic estimated Pulmonary Artery Systolic Pressure > 35 mmHg	1
E	**E**lder	Age > 60 years	1
F	**F**illing Pressure	Doppler Echocardiographic E/e' > 9	1
H_2FPEF score			Sum (0-9)



Description of the H2FPEF score and point allocations for each clinical characteristic (top box), with associated probability of having HFpEF based upon the total score as estimated from the model (lower box).(12)

5. Symptoms and clinical signs

Symptoms

- Dyspnea (general)
- Dyspnea on exertion and impaired exercise tolerance
- Orthopnea and paroxysmal nocturnal dyspnea

Clinical signs

- Peripheral edema (including ankle edema, lower leg edema, pitting edema)
- Basal crepitations