Fall risk prediction and validation in older adults

Leveraging electronic health records with machine learning

Dormosh, N.

Publication date
2023

Chapter

10

General discussion

This thesis has set out to leverage machine learning techniques and electronic health record (EHR) data to create algorithms and tools capable of quantifying fall risk in older adults. This provides a solid foundation for decision-making regarding effective interventions for reducing fall risk, particularly fall risk associated with medications. Specifically, we first examined existing prediction models constructed using EHR data for falls in older adults residing in the community. This involved systematically critiquing published models to gain insights into their strengths and limitations and formulate recommendations for future prediction models in this field. Subsequently, we developed and validated prediction models for falls utilizing routinely collected EHR data, both in primary care and in a hospital setting. Lastly, we investigated the potential application of natural language processing (NLP) and machine learning in predicting falls and in recognizing patterns of factors in clinical text notes contributing to the risk of falling.

In this chapter, we first provide a summary of the key findings. Subsequently, we concisely address these findings, and explore their significance and implications. Next, we highlight the strengths of the work presented in this thesis, acknowledge its limitations, suggest potential directions for future research, and ultimately conclude our discussion.

Summary of key findings

- Falls prediction models derived from routinely-collected EHR data had comparable performance to those developed using data from research cohorts, yet without the requirement for additional data collection.
- Methodological and statistical shortcomings were observed in existing prediction models for falls (notably inadequate internal validation, omission of calibration assessment, inadequate handling of missing data), and they lacked external validation.
- The prediction models for falls developed in this thesis turned out to contain a combination of predictors (e.g., demographics, medications, health conditions) aligning with the multifaceted nature of falls.
- The least absolute shrinkage and selection operator (Lasso) method was useful in combating overfitting, as demonstrated by consistent model performance during internal validation, and yielded parsimonious (few variables) prediction models in this thesis.
- The adoption of a methodological framework for external validation that uses the concept of differences between the training and test sets yielded valuable insights. Specifically, it showed that the performance of the developed model in the primary care setting was reproducible in an independent validation sample within the same setting.
- Unstructured clinical notes provided an additional viable means of fall risk estimation compared to existing traditional prediction models containing only struc-

tured clinical variables. The predictive performance slightly improved when the information from clinical notes complemented the clinical variables.

- The utilization of novel topic modelling techniques that leverage embedding approaches to capture semantics offered an efficient way for extracting interpretable latent topics from free text.
- Dynamic topic modelling provided useful insights when applied on general practitioners' longitudinal clinical notes to discover the trend of topics in relation to fall risk, and showed that an increase in healthcare needs appears to be a warning sign for future fall.

Main findings

Current state of falls prediction models

Many guidelines for falls prevention and interventions emphasize the use of fall-risk stratification tools to identify older adults at high risk of falling (1). These tools rely mainly on algorithms developed by clinical experts, considering handy factors like fall history and assessments of balance and gait such as the Timed Up and Go Test (TUG) (2). However, most of these algorithms lack validation, and the predictive ability of tests like the TUG is limited when used in isolation (3). Fall risk stratification can also be achieved by means of prediction models that offer individualized risk scores based on a large number of predictors (some of which may be known risk factors). Such prediction models can be derived from research cohorts or routinely-collected data. Prediction models for falls in community-dwelling older adults based on research cohorts (hereafter "cohort-based") were previously summarized and criticized in a systematic review (4). However, little is known about prediction models developed using routinely-collected data (hereafter "RCD-based"), with respect to performance, (internal/external) validation and transportability (i.e., generalizability of the results to other populations), particularly in comparison to cohort-based models. In **Chapter 2**, we included and reviewed 26 relevant studies, describing 30 prediction models (23 cohort-based and 7 RCD-based), and external validation of two existing models (one cohort-based model and one RCD-based). We found that discrimination performance was comparable between cohort-based and RCD-based models, with the respective areas under the receiver operating characteristic curves (AUC) ranging from 0.65 to 0.88 versus 0.71 to 0.81, for internally validated models. Importantly, the advantage of RCD-based models lies in their ability to achieve these results without the need for additional data collection efforts. However, the vast majority of studies on the development and validation of the models exhibited suboptimal reporting and were identified as having a high risk of bias. This high risk of bias was attributed to shortcomings in the statistical analysis and outcome determination, or inherent limitations associated with the use of routinely-collected data. This raises concerns about the reliability and generalizability of the models. Based on these shortcomings, we provided the following recommendations for researchers to improve prediction models for falls: (i) to adopt the recommended and well-established methodology when developing or validating prediction models (5, 6), including: adjusting for optimism, handling missing data, assessing model discrimination and calibration, and to identify/address potential biases using the PROBAST tool (7); (ii) to conduct more external validation studies for existing models in order to ensure their applicability in clinical practice; and (iii) to adhere to the TRIPOD statement (8) when reporting on prediction models for falls in order to enhance transparency and facilitate replication.

Fall prediction and validation

In **Chapter 3**, we developed and internally validated a prediction model for falls using primary care EHR data of 36,470 older individuals enlisted with 50 general practices in the Netherlands, of which 4,778 (13.1%) fell. The median AUC of the model was 0.705 (IQR 0.700–0.714) indicating fair discrimination performance, and the model demonstrated reasonable calibration, albeit with some overestimation of fall risk in a relatively small group of patients. The model included a combination of predictors, aligning with the multifaceted nature of falls, encompassing age and female sex, 5 health conditions, 2 medications, and the crucial component of fall history. To justify its clinical use and to evaluate the extent to which the results can be generalized, the model underwent external validation in **Chapter 4**. This validation was performed on an independent cohort of individuals, extracted from EHR data in a primary care setting. The validation cohort included a total of 39,342 older adults, of which 5,124 (13.4%) reported falls. The characteristics of the validation and the development cohorts were similar. The discriminative ability of the model as measured with the AUC was 0.690 (95% CI 0.686-0.698), and calibration performance was on parallel with our development study (**Chapter 3**). These results indicate that the model had good external validity, after reproducing its predictive performance in the validation cohort. As such, the model provides promising application in related primary care settings. In **Chapter 5**, we developed two prediction models to predict falls in a hospital setting, after 24 hours of admission. The first model (Model-without) utilized all potential predictors without indicators for missing values (dichotomous variables indicating the presence or absence of values). The second model (Model-with) combined both the potential predictors with the indicators. The AUCs of the Model-without and Model-with were 0.676 (95% CI: 0.646-0.707) and 0.695 (95% CI: 0.667-0.724), respectively, indicating fair discrimination. Our results showed that both models had fair discrimination and good calibration, where the model with indicators for missing values performed better. Like our findings in **Chapter 3**, both models incorporated a mix of predictor categories, capturing the multidimensional aspects of falls in older adults. The performance of our prediction model for falls in older adults residing in the community, as described in **Chapter 3** and **Chapter 4**, compares favorably and often surpasses models based on prospective research cohorts, as summarized in **Chapter 2** and in a previous review (4). Likewise, our RCD-based prediction models for inpatient falls, presented in **Chapter 5**, performed like the Johns Hopkins Fall Risk Assessment Tool (JHFRAT), a widely used tool in hospital settings (9–11). However, the advantage of our models is that they rely on readily available variables routinely collected during the medical care. This eliminates the need for extra data collection and decreases the workload on healthcare professionals, distinguishing our models from existing tools.

NLP and falls

In **Chapter 7**, we conducted an analysis to determine the predictive performance of utilizing general practitioners' (GPs) unstructured clinical notes for predicting future falls in older adults. We examined whether these notes alone or in combination with traditional structured clinical predictors could improve the performance. The approach was to deploy a topic modelling algorithm on the clinical notes in order to extract meaningful topics, and to use these topics as predictors for falls. We developed three logistic regression models using the least absolute shrinkage and selection operator: one using structured clinical variables (Baseline), one with topics extracted from unstructured clinical notes (Topic-based) and one by adding clinical variables to the extracted topics (Combi). AUCs and 95% confidence intervals of these models were 0.709 (0.700–0.719), 0.685 (0.676– 0.694) and 0.718 (0.708-0.727), respectively, and all the models showed good calibration. Regarding the topics inferred from the unstructured clinical notes, we found that institutionalisation, cognitive impairment, frailty, history of fractures or injuries were among the topics associated with higher fall risk and all are well-recognized in the literature (12–14). We also observed many other topics that were negatively associated with falls. These include, for instance, the topic of influenza vaccination, pre-travel health advice and vaccination, and topics focusing on cardiovascular risk management (CVRM) and diabetic care. The presence of these topics indicates that falls and fall-related information naturally exist in the unstructured clinical notes, suggesting that prediction models for falls based on the unstructured clinical notes provide an additional viable means of fall risk estimation compared to existing traditional prediction models containing structured clinical variables. Moreover, by integrating information extracted from the unstructured clinical notes with the structured clinical variables, we observed a small improvement in the predictive performance, albeit with limited clinical relevance.
In **Chapter 8**, we employed NLP to analyze and examine longitudinal clinical notes of GPs to discover potential factors contributing to future fall risk among older adults. Specifically, we applied dynamic topic modelling to extract latent topics from the clinical notes,

and to track these topics over time in relation to falls. We discovered 25 overarching topics serving as indicators of an imminent elevated risk of falls, such as “medications”, “renal care”, “family caregivers”, “referral/streamlining diagnostic pathways” and “hospital admission/discharge”. The majority of these topics demonstrated a clear pattern of increasing occurrence over time, suggesting potential warning signs before the occurrence of falls.

Medication-related falls

Certain medications are widely acknowledged as modifiable risk factors for falls. These medications, often referred to as Fall Risk Increasing Drugs (FRIDs), present an opportunity to mitigate the risk of falls when appropriately managed (15, 16). Moreover, they can be used as potential predictors when developing prediction models for falls. However, definitions of FRIDs vary between studies, and consensus on the exact grouping of these medications into standardized categories is lacking. In **Chapter 6**, we presented a case study in which we investigated the impact of describing medications at various levels of the hierarchy of the Anatomical Therapeutic Chemical (ATC) classification system (17). Our findings revealed that grouping medications at the second level of the ATC hierarchy, particularly the Pharmacological/Therapeutic level, can improve the predictive performance of models to predict falls in older adults. However, the optimal grouping level should be determined through practical experimentation. **Chapter 9** provided the work in this thesis in perspective of the SNOWDROP (SeNiors empOWred via big Data to joint-manage their medication-related Risk Of falling in Primary care) project. SNOWDROP aims to provide GPs and patients with tools to facilitate the shared-decision making process. We reported on the accomplished milestones and deliverables of the project. One of the key deliverables is the prediction model that provides a personalized fall risk estimation for older adults, which was developed in **Chapter 3** and validated in Chapter **Chapter 4**. Another important deliverable is the intervention consisting of (a) a personalized decision support system advising on reducing the patients’ medication-related fall risk, and (b) a patient portal. The personalized decision support system offers tailored guidance and support regarding the risks associated with each patient’s medication. To support this intervention, the patient portal serves as an online platform where patients can access their health information, including fall risk estimation and medication-related details, allowing them to prepare for the medication review consultation.

Significance and implications

Towards tailor-made predictions: personalizing the fall risk

As stated before, many guidelines on falls prevention and management rely on simple algorithms to stratify older adults into coarse risk groups (i.e., low or high) based on a handful of factors (1), including the recently introduced World Falls Guidelines (WFG) (18). Broad risk groups have their merits, particularly in large-scale screening initiatives to quickly identify individuals at higher falling risk and in contexts where data availability is limited. However, in clinical practice, optimal decisions should be based on a multitude of variables, resulting in finer grained and personalized risk predictions, in order to capture the complexity of each individual's risk profile. In this thesis, our prediction models provide a personalized risk estimation of falls by modelling the risk as probabilities, and they utilize multiple variables. This approach involves creating a scale ranging from 0 to 1 which enables a more nuanced decision-making process and tailored interventions. Such a scale becomes especially valuable for individuals who fall into the "gray zone" where their risk level is uncertain and requires a more precise evaluation. One of the advantages of probabilities is that they inherently provide error measures. For example, if the probability of a fall event is 0.2 and the decision is not to intervene, the probability of this being an error is, by definition, 0.2 for patients predicted not to fall. Similarly, probabilities can inform the clinician's actions. For instance, a higher probability of falling, such as 0.5, might prompt the clinician to take additional actions like conducting medication review or functional assessments. By considering this, clinicians can evaluate the potential consequences and make informed decisions based on the level of risk they are willing to accept according to the obtained individual's risk.

Desirability of personalized risk estimation becomes more evident in the realm of falls prevention, where shared-decision making is warranted and tools to estimate fall risk have limited discriminative ability (low AUC) (19, 20). Tools to perfectly predict falls in older adults do not exist and most likely will never exist. However, a calibrated prediction model that provides a continuous risk estimation for falls is informative. A predicted risk of 20% provided by a properly calibrated prediction model indicates that out of 100 older individuals with the same predicted risk, approximately 20 of them are likely to experience a fall. This information is particularly important in shared-decision making as it fosters a collaborative discussion between older adults and clinicians, empowering older adults to establish their own threshold (i.e., subjective boundary for taking preventive measures or actions). Consider an older adult with mobility issues who may be reluctant to use a walking aid due to concerns about stigma. This additional objective evidence of the individual's fall risk, perhaps with risk visualization, may help both the older adult and the clinician to explore the implications of the fall risk and the potential benefits of using a walking aid. Together, they can weigh the importance of addressing the fall risk against the concerns related to stigma, ensuring that the final decision aligns with the older adult's preferences and goals for maintaining their mobility and independence.

Our prediction model of falls in community-dwelling older adults have other advantages. The model relies on obtainable variables that are usually routinely collected during the medical care process, and do not require additional data collection such as functional assessments for gait and balance, which may take up to 20 minutes to complete (21). As such, our model can be potentially applied to estimate the fall risk in daily practice, without extra measures to collect data, saving the time and decreasing the burden on both healthcare professionals and older individuals. In addition, the model is transparent and can be easily interpreted, providing insight into the contribution of each variable in the model to the fall risk.

Although our prediction models can be useful to identify older adults at higher fall risk, it is essential for healthcare professionals to be mindful of several crucial considerations when using these models. It is important to note that the models have moderate ability to accurately identify older adults at higher fall risk, leading to missed opportunities for preventive interventions. Additionally, it is crucial to note that our models should not be interpreted as causal models, but rather as tools that quantify the fall risk and merely highlight the association between some predictors and falls. For example, the use of proton pump inhibitors (PPIs) was found to be an important predictor for falls in community-dwelling older adults. While PPIs have been associated with functional decline and with fracture risk which might increase the risk of fall-related injuries (22, 23), there is currently no evidence of a causal relationship between their use and falls. PPIs are likely surrogate of an underlying health condition or an indicator of frailty in older adults, who often have comorbidities requiring multiple medications.

From bench to bedside: integrating the models in EHRs

The use of computerized clinical decision support systems (CDSS) has revolutionized healthcare by assisting clinicians in making complex decisions. Since their introduction in the 1980s, the CDSS has undergone significant advancements. Nowadays, CDSSs can be integrated with EHR systems, facilitated by the rapid adoption of the EHR (24). Falls prevention and management are among the many domains that necessitate such intricate decision-making processes. Our prediction models, which offer personalized risk estimation for falls, can be valuable in the decision-making process, but their true benefits can only be realized when integrated within the EHR allowing for automatic provisioning of prediction scores, ideally coupled with recommendations. A previous study found that an important barrier of the adoption of prediction models is the lack of integration within EHR systems (25). Without integration, healthcare professionals need to manually transcribe variables into data fields prior to presenting score, which places considerable burden on them (25). The merit of using EHR data to develop and validate our models lies in the fact that these models rely on readily available and easily retrievable variables (26). Consequently, our models can be seamlessly integrated with a CDSS and implemented within the same EHR system.

However, implementing a prediction model in clinical practice poses considerable challenges. One significant challenge is the integration of the model within the existing clinical workflow (27–29). Healthcare professionals are already juggling multiple responsibilities and facing time constraints, so incorporating a new prediction model can disrupt their established routines. Another challenge is the acceptance and trust of healthcare professionals towards the prediction model. It is important to ensure that potential users, particularly clinicians, are well-informed about the model's underlying assumptions and predictors (e.g., which predictors are used and why) (30). The use of parametric regression models such as logistic regression can enhance such understanding, as their parameters provide a sense of how influential a predictor is on the predicted probability. However, sophisticated machine learning algorithms including many NLP prediction models are often perceived as black boxes by clinicians, where input data is processed through a hidden process to generate a prediction. Implementation of such models is challenging as clinicians may not fully comprehend the mechanism behind the predicted probabilities, potentially leading to incorrect intervention choices (30). Explainable Artificial Intelligence (XAI) attempts to provide insights into machine learning models' inner workings, helping clinicians understand and explain the reason behind the predicted scores generated by such models (31, 32). It is also important to engage clinicians in the early stages of model development, as this may help in pinpointing areas of concern within the prediction model, potentially leading to enhanced trust and facilitating improvements. An additional challenge lies in ensuring compliance with relevant regulations, particularly the General Data Protection Regulation (GDPR) in Europe. This poses an extra layer of complexity, especially for NLP prediction models that rely on free text data, which may potentially contain sensitive information. Implementation of such models require additional efforts to de-identify the free text and extra measures to improve data security in order to protect patients' rights. This becomes particularly crucial when the NLP prediction model and data are not housed within the EHR system of the clinician, requiring the transfer of text data to third parties for text processing and prediction score generation. To facilitate implementation, it is crucial to address these challenges, along with others that have been identified in previous studies (30, 33).

As mentioned above, the implementation of our falls prediction model within GPs' EHR systems, while ensuring its seamless integration into their workflow, presents a non-trivial challenge that requires careful consideration. This implementation requires an interdisciplinary initiative and a close collaboration between different stakeholders. The SNOWDROP project has been designed to accommodate the requirements for employing fall prediction models based on structured data, specifically the model described in **Chapter 3** in this thesis. The work in this thesis, except for **Chapter 9**, presents the first part of the SNOWDROP project in which we delved deep into prediction models and patterns in the data. The second part of the SNOWDROP project, described briefly in **Chapter 9**, involves extensive research and development activities from communication scientists, who effectively translated the prediction model and the CDSS into a practical application. Research and activities, in this phase, involved investigating barriers and facilitators of using a CDSS by clinicians (28). Additionally, interviews with older patients (results have not been published yet) and focus groups with GPs (29) were conducted to explore their needs and wishes for both the CDSS and the patient portal. The findings from these studies played a significant role in shaping the development and design choices for the initial prototypes of the CDSS and patient portal. For instance, based on the preferences expressed by the GPs during the focus groups, a visual representation of personalized fall risk using a gradient scale that ranged from bright green to dark red was incorporated within the CDSS. In addition, to ensure a seamless and uninterrupted workflow experience, the CDSS has been developed in such a way that a non-intrusive method is employed to notify GPs about potential fall risks. This involves the display of a notification, in the form of a colored button on the screen, allowing GPs to independently decide whether to access and review the advice provided, rather than relying on disruptive pop-ups. In the patient portal, a question prompt list was implemented to assist patients in preparing for consultations by indicating the questions or concerns they would like to discuss with their GP. Furthermore, a usability test was conducted involving GPs to evaluate the CDSS, and another usability test for older patients to assess the patient portal (results have not been published yet). Based on these usability tests, the prototypes were improved and a ready-to-implement versions of both systems were developed. Currently, the intervention of our prediction model integrated with a CDSS is being tested, in a randomized controlled trial, to investigate the impact of this intervention in comparison with usual care.

In the SNOWDROP project, the patient data, particularly the required predictors are sent anonymously via a secure channel to the ”cloud” where the NHGdoc system operates our prediction model in **Chapter 3** on these predictors to obtain a probability of fall. It then creates a form in which this probability is stated, along with medication-related recommendations to decrease fall risk. This form is accessible to the GP via the colored button that notifies the GP that there is a pending message. By clinking on the button, the GP will be engaged with the decision support system. The technology for the transfer of data, operating the prediction model, and engaging the GP is already in place. However, the deployment of our NLP based prediction models, based on free text, will require addressing the challenge of real-time text de-identification. The other technological elements of operating the NLP-based model, and engaging the GP with decision support are relatively easy to realize.

Pushing the boundaries: advancing methodology pertaining to falls prediction

The task of predicting falls in older adults is inherently challenging due to several reasons, and includes the complex interplay of both intrinsic (e.g., impaired balance, medication use, and chronic medical conditions) and extrinsic (e.g., environmental hazards, inadequate lighting, and slippery surfaces) factors. As a result, it is not unexpected to observe limited predictive capacity in existing prediction models. Our findings also underscore the importance of other aspects that could influence the performance of these prediction models, specifically relating to shortcomings in methodological and statistical analysis. Studies describing such models often lack internal and external validation, omit calibration assessment, and inappropriately handle missing data. In this thesis, we attempted the careful development and validation of prediction models, taking into account the aforementioned methodological limitations. We discuss below some important aspects that aim to advance the methodology related to falls prediction.

A common problem in the development of prediction models is model overfitting (6), where a prediction model learns patterns in the training data that do not generalize to unseen data, from the same population, leading to suboptimal model performance. One of the causes of overfitting is the use of too many variables in the prediction model (34). Variable selection is therefore important to reduce overfitting and hence to increase the model's performance on external data. This is particularly important when using EHR data for developing prediction models, as they offer a vast array of potential predictors (26). In order to anticipate the model's performance in new observations and to guide variable selection, internal validation should be employed when developing prediction models. Internal validation is the procedure of assessing model performance and the stability and the quality of the selected variables, on unseen observations from the same underlying population (5). Ideally, internal validation is performed using cross-validation or bootstrapping, and the whole model development strategy including variable selection and hyperparameter tuning should be repeated in this procedure (35, 36). In this thesis, our prediction models were developed using the Lasso (least absolute shrinkage and selection operator) procedure (37), a widely used regularization technique that allows for variable selection as part of fitting the model, ultimately decreasing overfitting. Also, the models underwent internal validation using cross-validation. As such, our models are more likely to maintain their performance when applied on new data having related observations.

External validation of prediction models is important, as their true value lies in their ability to deliver reliable performance in individuals beyond the dataset used for their development (38). External validation studies could be conducted to assess individuals who were recently treated (temporal validation), individuals from other geographical location (geographic validation), individuals from different clinical setting or even individuals from different target population (38). Consequently, the case mix in external validation studies can be different between the development and validation samples, encompassing differences in outcome rates, demographics, disease severity, healthcare settings, and target populations. Knowledge about the degree of the relatedness between the validation sample and the development sample is important to enhance our understanding and interpretation of the external validation results (39). Testing a prediction model on a validation sample that is similar to the development sample (e.g., temporal validation) indicates reproducibility testing, whereas testing on a validation sample that differs from the development sample in individual characteristics (e.g., different clinical setting) indicates transportability testing (39). To enhance our interpretation of our external validation study described in **Chapter 4**, we adopted the methodological framework proposed by Debray et al (39) which allowed us to interpret the results in terms of clinical reproducibility. We encourage researchers to incorporate this framework as it provides a robust methodological approach for enhancing the interpretation of external validation studies.

We stress the importance of correctly interpreting the risk of bias according to the PROBAST tool (7) when assessing prediction models. A low or high risk of bias does not necessarily imply perfect or imperfect models in itself, but rather indicates design and methodological weaknesses that impact the extent to which the model's results can be reproduced in unseen data. It is entirely possible for a model with a low risk of bias to perform poorly in an independent sample, just as a model with a high risk of bias may perform well. Irrespective of the degree of bias, conducting a proper external validation is paramount. While adhering to best practices during model development is crucial, only through rigorous external validation can we effectively rule out any potential bias and determine the model's true reliability and generalizability, especially when a prediction model relies on many proxy variables instead of causal ones. According to PROBAST, our prediction model developed in primary care setting should be by default rated with high risk of bias due to the use of EHR data, which are not explicitly designed for research purposes. However, the model showed good external validity when tested on an independent, related sample. This lends support for the model's applicability in external populations, though still related, particularly in the primary care setting in the Netherlands.

Most prediction model studies have missing data, but few addresses it statistically or discuss its implications (8). Missing data is particularly prominent in EHR data, where variables are collected based on their relevance to the patient (40), posing challenges for researchers when developing prediction models using EHR data. Various approaches can be employed to address missing data, spanning from simple exclusion of observations with missing values (complete-case analysis) to simple imputation methods like mean imputation, as well as more sophisticated techniques such as multiple imputation (41). In this thesis, we employed multiple imputation via chained equations (MICE)

(41, 42), which is considered the most appropriate and recommended method for handling missing data, when data is missing at random, as it leads to the least biased results (5, 7, 8). Additionally, we explored the potential benefits of utilizing indicators of missing values in our analysis (43, 44). The inclusion of these indicators showed improvements in the predictive performance. However, this approach should be exercised with caution as indicators of missing values have no specific clinical meaning but only a nonspecific one (45) and they also determine the prediction weight of other predictors. Furthermore, the generalizability of prediction models developed using such indicators is limited due to the potential variation in the underlying mechanisms that caused the missing values across different settings and their potential changes over time (39, 44).

The large amount of textual data in the EHR provides an opportunity to deepen our understanding of falls and fall related risk factors not explicitly coded in the structured data (46, 47). However, extracting such information manually poses significant challenges due to its labor-intensive nature and the requirement for prior medical knowledge to accurately capture the relevant details. In this thesis, we have demonstrated the usefulness of NLP, specifically through the utilization of topic modelling, to automatically extract valuable information in the form of topics. We utilized recent NLP approaches to perform topic modelling. In particular, we used Top2Vec (48) and BERTopic (49), which leverage several unsupervised machine learning algorithms to uncover latent topics or themes from text data. An important aspect of these algorithms is the utilization of advanced embedding techniques, such as doc2vec (50), and Large Language Models (LLM) like Bidirectional Encoder Representations from Transformers (BERT) (51), to map documents into dense vectors that capture the contextual understanding and the semantic relationships between documents. As such, they extract interpretable and coherent topics to a large extent. This approach offers researchers an efficient means to obtain tentative initial evidence, fostering a preliminary understanding of the factors associated with fall risks. Ultimately, this information can inform subsequent research endeavours, interventions, and the development of preventive strategies to mitigate fall risks.

Diving deeper: unlocking clinical notes

Among the meaningful topics extracted from GPs' clinical notes (**Chapter 7**), harbingers of future falls in community-dwelling older adults included the topics cognitive impairment, frailty, prior fractures or injuries. Frailty and prior injuries are among the factors considered in the WFG's algorithm for fall risk stratification. On the other hand, cognitive impairment has not been included before in most of the risk stratification tools provided by the existing guidelines for falls, and specific recommendations for older adults with cognitive impairment were scarce (1), although it is a widely recognized risk factor for falls (52). Interestingly, cognitive impairment remained a strong predictor for falls when we complemented the topics extracted from the clinical notes with the structured clinical variables. This finding highlights the importance of considering cognitive function in fall risk estimation and provides some support to include it as a predictor in fall-risk stratification tools.

Equally important is gaining insight into the topics negatively associated with falls. Though not necessarily causal, these topics may shed light on preventive and protective measures that can be implemented to reduce the occurrence of falls. For example, the topic on influenza vaccination was found to be protective for falls. This finding emphasizes the potential role of influenza vaccination in preventing influenza symptoms, such as fatigue, dizziness, and abnormal gait, all of which are recognized as factors that make individuals more prone to falling (14). Another example is the presence of various topics centred on cardiovascular risk management (CVRM) and diabetic care. These topics were found to be negatively associated with falls, indicating their potential protective influence. This finding underscores the importance of addressing underlying conditions that can contribute to falls, as explicitly emphasized in the WFG (18).

Risk factors for falls are not static but change over time. It is therefore important for clinicians to comprehend the entire trajectory of patient care including the trend and changes of the risk factors over time. In this thesis, our NLP trend analysis of longitudinal clinical notes from GPs revealed numerous topics implying risk factors for falls, demonstrating a consistent upward trend leading up to fall incidents. We found some topics that pertain to a specific medical domain such as “medications” and “renal care”, implying identifiable risk factor for falls (e.g., FRIDs, chronic kidney disease). An increase in the prominence of these topics over time, as reflected in the number of clinical notes, may indicate the emergence or deterioration of medical conditions, which could foreshadow falls. Likewise, many other topics such as “hospital admission/discharge”, “referral/streamlining diagnostic pathways”, “communication between healthcare providers” indicate an increase in healthcare utilization and care needs. These topics suggest a broad spectrum of underlying health conditions or complications involving specialized care, ongoing monitoring or progressing of functional limitations or increased care needs, which can result in a higher risk of falls. This could be seen as a signal and a warning sign of a future fall, a crisis that could possibly be avoided if measures (multifactorial assessment) are timely taken. Clinicians should proactively identify any potential rise in an individual’s healthcare needs. It is crucial for them to conduct thorough assessments that address the underlying causes, e.g., identify underlying new or worsening conditions and implement suitable interventions to decrease the fall risk. Furthermore, researchers involved in the development of falls prediction models may consider taking into account the change of predictors over time, as this could enhance the accuracy of the predictions.

Strengths and limitations

Harnessing EHR data to develop prediction models for clinical outcomes provide opportunities and challenges (26), yet the specific nuances and complexities associated with predicting falls using EHR data have received limited attention in the current literature. A notable strength of our work lies in our systematic approach to summarize and critically evaluate existing prediction models for falls based on EHR data. By gaining valuable insights into the current situation, we were able to foresee and steer clear of the pitfalls encountered in previous models during the development of our prediction models described in this thesis. We adopted a rigorous approach in the development and validation of our prediction models, guided by best methodological practices, and transparently reported our results according to well-established reporting standards. Moreover, our work is based on a large sample size comprising representative older adults. Another important strength is the conduction of an external validation of our primary care falls prediction model, which was found to be lacking in existing models. Furthermore, our work stands out by integrating innovative NLP techniques, showcasing their promising potential in predicting falls and uncovering associated risk factors, while also capturing their temporal trends. This approach allowed us to gain valuable insights into the dynamic nature of falls, facilitating a deeper understanding of the underlying factors and their evolution over time. Lastly, a major strength lies in the interdisciplinary nature of the SNOWDROP project, which bridges the fields of medical, data, and communication sciences to achieve its objectives. This interdisciplinary collaboration effectively addresses the complexity of fall risk management in older adults by empowering both GPs and older individuals with reliable tools. This empowerment facilitates more informed shared decision-making and ultimately contributes to a reduction in the risk of falls.

Alongside the abovementioned strengths, it is crucial to recognize and address the limitations, primarily those inherent to the EHR data and retrospective analysis. Although the fall rates found in our studies were higher than other studies conducted using EHR data, they remain lower than those reported in prospective cohort studies. Unlike prospective cohort studies in which falls can be consistently tracked using fall diaries, falls in our studies were encoded in the narrative clinical notes. Not all falls are necessarily documented by GPs or reported by older adults, potentially leading to an underrepresentation of fall incidents. Older adults frequently fail to report falls, either due to recall bias, a reluctance stemming from the stigma attached to falling or their perception that the falls were insignificant enough to warrant reporting. Consequently, it is likely that the falls included in our studies were primarily those that required medical attention (53). However, falls in our studies may better align with real clinical practice, where a significant proportion of non-injurious falls tend to go unreported. As a result, our primary care prediction model accurately captures the characteristics of falls in real-world settings, enhancing the generalizability of its results, as demonstrated during the external validation. A further limitation specific to our primary care falls prediction model pertains to the exclusion of laboratory measurements in the analysis. This decision was made due to the extensive number of missing values, as a substantial number of individuals did not have these measurements taken. Nevertheless, our exploratory experiments did not show any improvement in the predictive performance when we incorporated indicators for missing values, nor the use of an algorithm such as the eXtreme Gradient Boosting (XGBoost) (54), which has the capability to handle missing values. In addition, we did not investigate the predictive value of certain potential predictors, including mobility, gait, and environmental factors, as such data are not routinely collected in the EHR. Nonetheless, we hypothesize that some of the other predictors (e.g., osteoarthritis) may have served as surrogates for these predictors. A potential limitation of the majority of the studies described in this thesis, and in particular studies focused on NLP, is the reliance on EHR data from a single healthcare system, in a unique setting where GPs play the role of gatekeepers. It should be acknowledged that documentation styles and clinical practices vary among healthcare providers, regions and countries. Therefore, the extent to which our findings generalize to other contexts remains to be established. For example, our prediction models developed using the clinical notes of the GPs are limited to the Dutch language but can be trained using our development strategy to make them applicable for another languages.

Future perspectives

Aside from the future work described above, there are additional avenues for further research and exploration. As the aging population continues to grow, it becomes increasingly crucial to develop reliable and more accurate prediction models to identifying individuals at risk of falling to prevent fall-related consequences and improve their well-being. In this thesis, we developed our falls prediction models using a machine learning approach that prioritize interpretability and simplicity. Specifically, we applied regularized logistic regression, which assumes a linear relationship between the independent and dependant variables. Moreover, falls are complex events influenced by multiple interacting factors. Therefore, future research could explore alternative machine learning algorithms, such as random forests, XGBoost, and neural networks, that can consider these intricate relationships and handle the non-linear relationships between independent and dependent variables. When these latter models would be demonstrated to exhibit significantly better performance than simple algorithms, clinicians may consider prioritizing performance over interpretability. Future research is required to determine whether the inclusion of the underexplored predictors, identified in the work of this thesis, such as lab measurements, vital signs and those in textual data could improve the performance.

Prediction models should only be considered for practical implementation when they underwent external validation (8, 55, 56). Without such validation, it is important to approach the model's performance with scepticism. We conducted external validation for our prediction model for falls in community-dwelling older adults in order to justify its applicability in primary care setting in the Netherlands. Prior to deploying our model in a particular setting or in another location, researchers should carry out a validation study specific to that location, and to continuously monitor the model's performance and make updates, especially when calibration becomes challenging (57). Likewise, future work is needed to externally validate our prediction models developed in a hospital setting, especially the one developed using missing indicators for missing values, as it might not generalize to other hospitals or other settings where the underlying mechanism that produced the missing values may differ (44).

A unique aspect of the work in this thesis is the use of NLP to predict future falls and to discover the trend of fall risk factors over time in relation to fall risk. We opted to represent the clinical notes as a collection of topics that can be interpreted by clinicians to understand topics contributing to fall risk. More work will need to be done to investigate different text representation approaches, and the impact of incorporating temporal information extracted from the clinical notes on the predictive performance. In addition, a natural progression of this work is to analyze several hypotheses raised in this thesis with regard to the temporal trend of certain topics pertaining to fall risk factors, such as those revolving around medications, certain health conditions, healthcare utilization and frailty. Conducting such studies can enhance our understanding of fall risk factors and potentially guide the development of effective falls prevention and intervention strategies.

Finally, the true value of our intervention, encompassing the prediction model for falls and a CDSS, can be fully realized when a thorough evaluation is conducted to assess its impact on clinical healthcare following implementation. This evaluation plays a vital role in addressing essential questions: How does the intervention influence clinical decision-making? And how do these decisions subsequently impact the health outcomes of patients and the community at large? The findings from these investigations have the potential to shape policy and public health initiatives, facilitating the expansion of the intervention on a national scale.

Conclusion

This thesis aimed to develop, validate and improve prediction models for falls in older adults by utilizing EHR data and harnessing the power of machine learning, including NLP of free-text clinical notes. Prediction models based on EHR data provide an individualized risk estimation for falls and can be seamlessly integrated with a CDSS and implemented in EHR platforms due to their reliance on routinely-collected variables. NLP of free-text clinical notes demonstrated encouraging results to predict future falls, particularly when information extracted using NLP is augmented with the structured clinical variables, and in discovering potential information hidden within the clinical notes. By implementing the prediction models provided in this thesis, along with a CDSS within EHR systems, clinicians can easily access the predictions and recommendations generated by the CDSS. This integration eliminates the requirement for additional data collection and empowers clinicians to make informed decisions and implement customized interventions to effectively minimize falls and mitigate fall-related injuries.

References

[1] Manuel M Montero-Odasso, Nellie Kamkar, Frederico Pieruccini-Faria, Abdelhady Osman, Yanina Sarquis-Adamson, Jacqueline Close, David B Hogan, Susan Winifred Hunter, Rose Anne Kenny, Lewis A Lipsitz, et al. Evaluation of clinical practice guidelines on fall prevention and management for older adults: a systematic review. *JAMA network open*, 4(12):e2138911–e2138911, 2021.

[2] Anne Shumway-Cook, Sandy Brauer, and Marjorie Woollacott. Predicting the probability for falls in community-dwelling older adults using the timed up & go test. *Physical therapy*, 80(9):896–903, 2000.

[3] Gotaro Kojima, Tahir Masud, Denise Kendrick, Richard Morris, Sheena Gawler, Jonathan Treml, and Steve Iliffe. Does the timed up and go test predict future falls among british community-dwelling older people? prospective cohort study nested within a randomised controlled trial. *BMC geriatrics*, 15(1):1–7, 2015.

[4] G.V. Gade, M.G. Jørgensen, and Ryg J. Predicting falls in community-dwelling older adults: a systematic review of prognostic models. *BMJ Open*, 11:044170, 2021.

[5] E.W. Steyerberg and Y. Vergouwe. Towards better clinical prediction models: Seven steps for development and an abcd for validation. *Eur. Heart J*, 2014.

[6] Ewout W. Steyerberg. *Clinical Prediction Models*. Springer International Publishing, 2019.

[7] R.F. Wolff, K.G.M. Moons, R.D. Riley, P.F. Whiting, M. Westwood, G.S. Collins, J.B. Reitsma, J. Kleijnen, and S. Mallett. Probast: A tool to assess the risk of bias and applicability of prediction model studies. *Ann. Intern. Med*, 170:51–58, 2019.

[8] K.G.M. Moons, D.G. Altman, J.B. Reitsma, J.P.A. Ioannidis, P. Macaskill, E.W. Steyerberg, A.J. Vickers, D.F. Ransohoff, and G.S. Collins. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): Explanation and elaboration. *Ann. Intern. Med*, 162:1– 73, 2015.

[9] K.S. Kim, J.A. Kim, and Y.K. Choi. A comparative study on the validity of fall risk assessment scales in korean hospitals. *Asian Nurs Res (Korean Soc Nurs Sci*, 5(1):1317 11 60011–, 2011.

[10] S.S. Poe, P.B. Dawson, and M. Cvach. The johns hopkins fall risk assessment tool: A study of reliability and validity. *J Nurs Care Qual*, 33(1):10–19, 2018.

[11] W.D. Klinkenberg and P. Potter. Validity of the johns hopkins fall risk assessment tool for predicting falls on inpatient medicine services. *J Nurs Care Qual*, 32(2):108–113, 2017.

[12] Marta Aranda-Gallardo, Jose M Morales-Asencio, Margarita Enriquez de Luna-Rodriguez, Maria J Vazquez-Blanco, Juan C Morilla-Herrera, Francisco Rivas-Ruiz, Juan C Toribio-Montero, and Jose C Canca-Sanchez. Characteristics, consequences and prevention of falls in institutionalised older adults in the province of malaga (spain): a prospective, cohort, multicentre study. *BMJ Open*, 8(2):e020039, February 2018.

[13] Mei-Ling Ge, Eleanor M Simonsick, Bi-Rong Dong, Judith D Kasper, and Qian-Li Xue. Frailty, with or without cognitive impairment, is a strong predictor of recurrent falls in a US population-representative sample of older adults. *The Journals of Gerontology: Series A*, 76(11):e354–e360, mar 2021.

[14] Anne Felicia Ambrose, Geet Paul, and Jeffrey M. Hausdorff. Risk factors for falls among older adults: A review of the literature. *Maturitas*, 75(1):51–61, may 2013.

[15] L. J. Seppala, , N. van der Velde, T. Masud, H. Blain, M. Petrovic, T. J. van der Cammen, K. Szczerbińska, S. Hartikainen, R. A. Kenny, J. Ryg, P. Eklund, E. Topinková, A. Mair, L. Laflamme, H. Thaler, G. Bahat, M. Gutiérrez-Valencia, MA Caballero-Mora, F. Landi, M. H. Emmelot-Vonk, A. Cherubini, J. P. Baeyens, A. Correa-Pérez, A. Gudmundsson, A. Marengoni, D. O'Mahony, N. Parekh, F. E. Pisa, C. Rajkumar, M. Wehling, and G. Ziere and. EuGMS task and finish group on fall-risk-increasing drugs (FRIDs): Position on knowledge dissemination, management, and future research. *Drugs and Aging*, 36(4):299–307, feb 2019.

[16] L.J. Seppala, M. Petrovic, and J. Ryg. Stoppfall (screening tool of older persons prescriptions in older adults with high fall risk): a delphi study by the eugms task and finish group on fall risk-increasing drugs. *Age Ageing*, 50(4):1189–1199, 2021.

[17] WHO Collaborating Centre for Drug Statistics Methodology Norwegian Institute of Public Health. Guidelines for atc classification and ddd assignment, 2022. norwegian institute of public health (2021.

[18] Manuel Montero-Odasso, Nathalie van der Velde, Finbarr C Martin, Mirko Petrovic, Maw Pin Tan, Jesper Ryg, Sara Aguilar-Navarro, Neil B Alexander, Clemens Becker, Hubert Blain, et al. World guidelines for falls prevention and management for older adults: a global initiative. *Age and ageing*, 51(9):afac205, 2022.

[19] Laure Wynants, , Maarten van Smeden, David J. McLernon, Dirk Timmerman, Ewout W. Steyerberg, and Ben Van Calster. Three myths about risk thresholds for prediction models. *BMC Medicine*, 17(1), October 2019.

[20] Anne AH de Hond, Ewout W Steyerberg, and Ben van Calster. Interpreting area under the receiver operating characteristic curve. *The Lancet Digital Health*, 4(12):e853–e855, 2022.

[21] P. Palumbo, C. Becker, S. Bandinelli, and L. Chiari. Simulating the effects of a clinical guidelines screening algorithm for fall risk in community dwelling older adults. *Aging Clin Exp Res*, 31:1069–76, 2019.

[22] L.J. Seppala, E.M.M. Glind, and J.G. Daams. Fall-risk-increasing drugs: a systematic review and meta-analysis: Iii. others. *Others. J Am Med Dir Assoc*, 19(4), 2018.

[23] K. Lapumnuaypol, C. Thongprayoon, K. Wijarnpreecha, A. Tiu, and W. Cheungpasitporn. Risk of fall in patients taking proton pump inhibitors: a meta-analysis. *QJM*, 112(2):115–121, 2019.

[24] Reed T Sutton, David Pincock, Daniel C Baumgart, Daniel C Sadowski, Richard N Fedorak, and Karen I Kroeker. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine*, 3(1):17, 2020.

[25] Videha Sharma, Ibrahim Ali, Sabine van der Veer, Glen Martin, John Ainsworth, and Titus Augustine. Adoption of clinical risk prediction tools is limited by a lack of integration with electronic health records. *BMJ Health and Care Informatics*, 28(1):e100253, feb 2021.

[26] B.A. Goldstein, A.M. Navar, M.J. Pencina, and J.P.A. Ioannidis. Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. *J Am Med Informatics Assoc*, 24(1):198–208, 2017.

[27] Saif Khairat, David Marc, William Crosby, and Ali Al Sanousi. Reasons for physicians not adopting clinical decision support systems: Critical analysis. *JMIR Medical Informatics*, 6(2):e24, April 2018.

[28] L. Westerbeek, K.J. Ploegmakers, G.J. Bruijn, A.J. Linn, J.C.M. Weert, J.G. Daams, N. Velde, H.C. Weert, A. Abu-Hanna, and S. Medlock. Barriers and facilitators influencing medication-related cdss acceptance according to clinicians: A systematic review. *International Journal of Medical Informatics*, 152:104506, 2021.

[29] L. Westerbeek, G.J. Bruijn, H.C. Weert, A. Abu-Hanna, S. Medlock, and J.C.M. Weert. General practitioners' needs and wishes for clinical decision support systems: A focus group study. *International Journal of Medical Informatics*, 168:104901, 2022.

[30] Teus H Kappen, Wilton A van Klei, Leo van Wolfswinkel, Cor J Kalkman, Yvonne Vergouwe, and Karel GM Moons. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagnostic and prognostic research*, 2(1):1–11, 2018.

[31] Ribana Roscher, Bastian Bohn, Marco F. Duarte, and Jochen Garcke. Explainable machine learning for scientific insights and discoveries. *IEEE Access*, 8:42200–42216, 2020.

[32] Vaishak Belle and Ioannis Papantonis. Principles and practice of explainable machine learning. *Frontiers in Big Data*, 4, July 2021.

[33] Joshua Watson, Carolyn A Hutyra, Shayna M Clancy, Anisha Chandiramani, Armando Bedoya, Kumar Ilangovan, Nancy Nderitu, and Eric G Poon. Overcoming barriers to the adoption and implementation of predictive modeling and machine learning in clinical care: what can we learn from US academic medical centers? *JAMIA Open*, 3(2):167–172, 2020.

[34] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[35] Ewout W Steyerberg and Frank E Harrell. Prediction models need appropriate internal, internal–external, and external validation. *Journal of clinical epidemiology*, 69:245–247, 2016.

[36] Ewout W Steyerberg. Validation in prediction research: the waste by data splitting. *Journal of clinical epidemiology*, 103:131–133, 2018.

[37] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, January 1996.

[38] Chava L Ramspek, Kitty J Jager, Friedo W Dekker, Carmine Zoccali, and Merel van Diepen. External validation of prognostic models: what, why, how, when and where? *Clinical Kidney Journal*, 14(1):49–58, November 2020.

[39] T.P.A. Debray, Y. Vergouwe, H. Koffijberg, D. Nieboer, E.W. Steyerberg, and K.G.M. Moons. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*, 68(3):279–289, 2015.

[40] Irene Petersen, Catherine A Welch, Irwin Nazareth, Kate Walters, Louise Marston, Richard W Morris, James R Carpenter, Tim P Morris, and Tra My Pham. Health indicator recording in uk primary care electronic health records: key implications for handling missing data. *Clinical epidemiology*, pages 157–167, 2019.

[41] Stef van Buuren. *Flexible imputation of missing data*. CRC Press, 2018.

[42] Stef Van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 45:1–67, 2011.

[43] Amelia LM Tan, Emily J Getzen, Meghan R Hutch, Zachary H Strasser, Alba Gutiérrez-Sacristán, Trang T Le, Arianna Dagliati, Michele Morris, David A Hanauer, Bertrand Moal, et al. Informative missingness: What

can we learn from patterns in missing laboratory data in the electronic health record? *Journal of Biomedical Informatics*, 139:104306, 2023.

[44] M. Sperrin, G.P. Martin, R. Sisk, and N. Peek. Missing data should be handled differently for prediction than for description or causal explanation. *J Clin Epidemiol*, 125:183–187, 2020.

[45] Heijden GJMG, T.Donders AR, Stijnen T, and Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: A clinical example. *J Clin Epidemiol*, 59(10):1102–1109, 2006.

[46] Ragnhildur I Bjarnadottir and Robert J Lucero. What can we learn about fall risk factors from ehr nursing notes? a text mining study. *eGEMs*, 6(1), 2018.

[47] H. Kharrazi, L.J. Anzaldi, and Hernandez L. The value of unstructured electronic health record data in geriatric syndrome case identification. *J Am Geriatr Soc*, 66:1499–507, 2018.

[48] Dimo Angelov. Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*, 2020.

[49] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.

[50] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1188–1196, Bejing, China, 22–24 Jun 2014. PMLR.

[51] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018.

[52] Susan W. Muir, Karen Gopaul, and Manuel M. Montero Odasso. The role of cognitive impairment in fall risk among older adults: a systematic review and meta-analysis. *Age and Ageing*, 41(3):299–308, 2012.

[53] J.A. Stevens, M.F. Ballesteros, K.A. Mack, R.A. Rudd, E. DeCaro, and G. Adler. Gender differences in seeking care for falls in the aged medicare population. *Am J Prev Med*, 43(1):59–62, 2012.

[54] T. Chen and C. Guestrin. Xgboost: a scalable tree boosting system. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 785–794, San Francisco, California, USA, 2016.

[55] D.G. Altman, Y. Vergouwe, P. Royston, and K.G.M. Moons. Prognosis and prognostic research: validating a prognostic model. *BMJ*, 338:1432–1435, 2009.

[56] E.W. Steyerberg and F.E. Harrell. Prediction models need appropriate internal, internal-external, and external validation. *J. Clin. Epidemiol*, 69:245–247, 2016.

[57] B. Calster, E.W. Steyerberg, L. Wynants, and M. Smeden. There is no such thing as a validated prediction model. *BMC Med*, 21:70, 2023.