NIPH
Norwegian Institute of Public Health

REPORT

2021

Aims and strategy for the implementation of machine learning in evidence synthesis in the Cluster for Reviews and Health Technology Assessments for 2021-2022

Title	Aims and strategy for the implementation of machine learning in evidence synthesis in the Cluster for Reviews and Health Technology Assessments for 2021-2022
Norwegian title	Mål og strategi for implementering av maskinlæring i kunnskapsoppsummeringer i klynge for vurdering av tiltak 2021-2022
Institution	Norwegian Institute of Public Health Camilla Stoltenberg, direktør
Authors	Ashley Elizabeth Muller, project leader, Folkehelseinstituttet Heather Ames, Folkehelseinstituttet Jan Himmels, Folkehelseinstituttet Patricia Jacobsen Jardim, Folkehelseinstituttet Lien Nguyen, Folkehelseinstituttet Christopher Rose, Folkehelseinstituttet Stijn Van De Velde, Folkehelseinstituttet
ISBN	978-82-8406-234-1
Type of report	Report (Rapport)
No. of pages	17
Client	Norwegian Inistitute of Public Health
Subject heading (MeSH)	Machine learning; systematic review; automation; technological assessment, biomedical; technological assessment, health; unsupervised machine learning; supervised machine learning; deep learning
Citation	Muller AE, Ames H, Himmels J, Jardim PJ, Nguyen L, Rose C, Van De Velde S. Mål og strategi for implementering av maskinlæring i kunnskapsoppsummeringer i klynge for vurdering av tiltak 2021-2022 [Aims and strategy for the implementation of machine learning in evidence synthesis in the Cluster for Reviews and Health Technology Assessments for 2021-2022]. Report – 2021. Oslo: Norwegian Institute of Public Health.

Table of contents

Key messages

In 2020-2021, a team in the Cluster for Reviews and Health Technology Assessments, Division for Health Services at the Norwegian Institute of Public Health (NIPH) ran a project on machine learning (ML) related to the conduct of evidence syntheses. Part of the work involved creating a vision and proposals for expanding ML activities in 2021-2022.

This report describes the team's suggestion for a strategic approach to meeting the continued need for innovation, evaluation, and implementation of ML for health technology assessments, systematic reviews, and other evidence syntheses. We propose a vision and goals, and a novel and flexible team structure. We divide activities into innovation, evaluation, and implementation, and present a risk assessment to inform the roll-out of a future team working on ML activities.

Title:

Aims and strategy for the implementation of machine learning in evidence synthesis in the Cluster for Reviews and Health Technology Assessments for 2021-2022

Publisher:

The Norwegian Institute of Public Health conducted the project based on an initiative by the Cluster for Reviews and Health Technology Assessments, Division for Health Services at the NIPH

Type of publication:
Report

Activity timeline:
Dec 2020 - June 2021

Hovedbudskap

Et lag i Klynge for vurdering av tiltak, Område for helsetjenester ved Folkehelseinstituttet undersøkte i 2020-2021 bruken av maskinlæring i kunnskapsoppsummeringer. En del av arbeidet var å utforme et overordnet mål og en strategi for å kunne oppskalere maskinlæring i framtida.

Denne rapporten beskriver lagets forslag til en strategisk tilnærming for å møte behovet for ytterligere bruk av maskinlæring i metodevurderinger, systematiske oversikter, og andre typer kunnskapsoppsummeringer. Vi forslår en visjon og flere mål, samt en ny og fleksibel lagstruktur. Vi beskriver nøkkelaktiviteter når det gjelder *innovasjon*, *evaluering*, og *implementering*, og presenterer en risikovurdering som kan støtte framtidig oppstart av et nytt maskinlæringslag.

Tittel:

Mål og strategi for implementering av maskinlæring i kunnskapsoppsummeringer i klynge for vurdering av tiltak 2021- 2022

Hvem står bak denne publikasjonen?

Folkehelseinstituttet utførte studien basert på et initiativ fra klynge for vurdering av tiltak, område for helsetjenster i FHI

Type publikasjon:

Rapport

Tidsperiode for prosjektet:

Des 2020 - Juni 2021

Preface

In 2020-2021, a team in the Cluster for Reviews and Health Technology Assessments, Division for Health Services at the Norwegian Institute of Public Health (NIPH) ran a project on machine learning related to the conduct of evidence syntheses. The team tested and documented pros and cons of using machine learning in various phases of the conduct of evidence syntheses, and built employees' competence in using machine learning. This report describes their aims and proposals for expanding machine learning activities in 2021-2022.

The report is relevant for researchers and managers interested in implementing machine learning in their evidence syntheses. It is particularly relevant for evidence synthesis environments that do not have machine learning specialists.

Financing
The work carried out by the machine learning team was self-initiated and financed by the Cluster for Reviews and Health Technology Assessments, Division for Health Services at the NIPH.

Team members
Project leader: Ashley Elizabeth Muller
Team members: Heather Ames, Jan Himmels, Patricia Jacobsen Jardim, Lien Nguyen, Christopher Rose, Stijn Van De Velde

Conflicts of interest
All authors declare they have no conflicts of interest.

Kåre Birger Hagen	Rigmor C Berg	Ashley E. Muller
Research director	*Department director*	*Project leader*

Background

There is an increasing demand from systematic review users/commissioners for high-quality reviews delivered faster, with greater efficiency, and at lower cost. There is also a growing societal need for high-quality, understandable, and accessible knowledge. Rapid developments in the types of data and advanced methods available are opening opportunities to increase efficiency and speed without compromising quality. One critical opportunity is **machine learning (ML)**. In a separate report, available on the NIPH website, we describe ML work undertaken by a ML team at the NIPH, project results and lessons learned. Our perspective is:

The Problem

HTV produces review products according to the methodological gold standard, but this is unsustainable. Commissioners are increasingly demanding evidence be summarized faster to use it to make decisions about health and welfare intervention implementation. Rapid reviews are a common response to short deadlines, but these often involve methodological shortcuts that can weaken our confidence in their findings.

The Solution

A one-year machine learning (ML) team that works with a continuous cycle of innovation, evaluation, and implementation activities to integrate the most effective ML functions into existing workflows, increase speed of evidence synthesis production, significantly scale up HTV's capacity to use ML, and maintain high quality of products.

The Potential

Full integration of ML and corresponding workflow optimizations - including data-sharing and automation - into all future reviews, with ML functions matched to review types, questions, deadlines, and sizes. Reviews are produced with fewer resources and in less time, with similarly high quality.

Situating ML within existing institutional priorities

NIPH recently revised its five-year strategic priorities. Becoming a leader in ML, big data, and modelling within the field of public health is now included as a priority. In addition, the Division for Health Services aims to become a *leader in automation and digi-*

talisation of work processes, while the Cluster for Reviews and Health Technology Assessments (HTV) is specifically mandated to implement new digital tools and continue improving workflows in order to produce evidence syntheses faster and more accessibly.

Using a ML team to meet institutional priorities

The machine learning team's 2020-2021 report, , available on the NIPH website, details the work undertaken by the first ML team, project results and lessons learned. There remains outstanding potential that we recommend be explored.

The 2020-2021 ML team focused on implementation of ML functions within HTV's projects. We believe it is worthwhile to continue this work. First, prioritizing projects within HTV; second, sharing successes to other groups within the Division for Health Services working on evidence syntheses and health technology assessments; and third, focusing on knowledge transfer and novel applications of ML across HTV and beyond. Based on our recent mapping of ML activities, we have identified ML activities and innovation outside of evidence synthesis occurring across the NIPH. If coordinated, these activities have the potential for significant knowledge transfer and more efficient capacity-building.

Aims and strategy

Goals for 2021-2022

Our ‘vision’ for 2021-2022:

- HT is an implementation leader in applying ML functions and adapting workflows to produce effective, high-quality evidence syntheses.
- The user spectrum, from researchers to commissioners, accepts ML functions as usual practice within evidence synthesis.

Our goals for 2021-2022:

1. To consolidate knowledge of and support for ML functions. Researchers and librarians understand the basics of ML, “why”, “when”, and “how” to use various functions in evidence syntheses.
2. To increase ML capacity from HTV to HT, informed by institution-wide knowledge and resources.
3. To identify and test new developments in the field of ML to assess how they could improve our workflows and products.
4. To scale up the number of teams independently using ML functions as part of their projects. By June 2022, all teams in HTV have the necessary knowledge and confidence to implement at least one ML function in their projects.

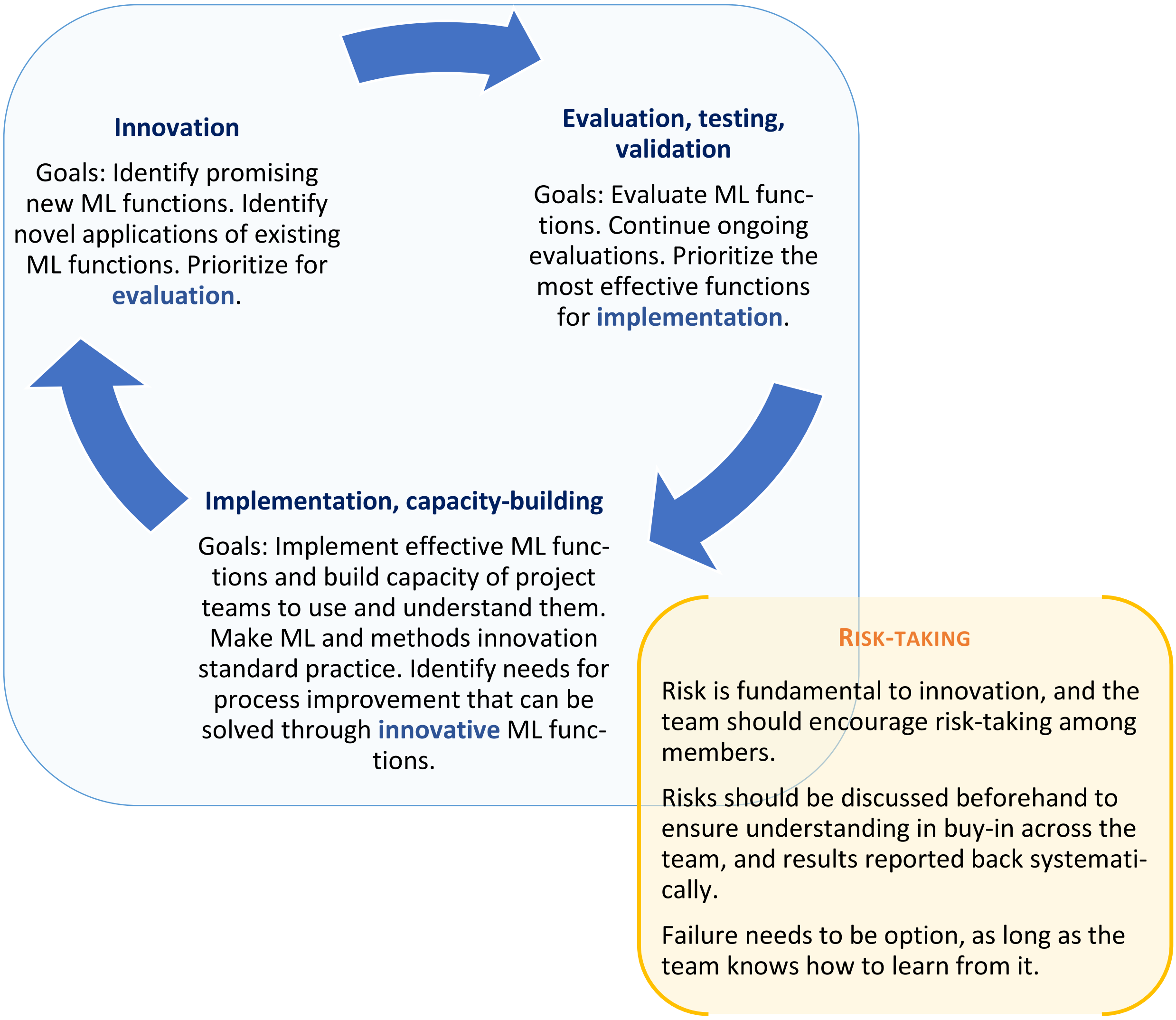
Timeframe

This strategy suggests a one-year term. The first ML team had a duration of six months. While this was sufficient for some innovation and testing, it was insufficient for longer-term goals such as HTV capacity-building and testing of ML functions used by less common evidence syntheses. There was also a five-week sunk cost in bringing all team members up to date with ML functions. A one-year term also ensures a sufficient pool of projects within which to work, as most HTV projects take between six and twelve months.

Activities

We envision the ML team continuing its pattern of activities of innovation (identifying promising new functions and applications to evaluate), evaluation (prioritizing the most effective functions to be used by projects), and implementation (teaching project teams to use ML functions independently). In this iteration, we have higher ambitions and will have a continual focus on scaling up activities and knowledge transfer, i.e. collaboration with and learning from other groups internally and externally (figure 1).

Figure 1: *Proposed activities for a future ML team*



Innovation activities

Horizon scanning and team learning through external collaboration

At present, the software and technical development side of evidence synthesis-specific ML activities originate predominantly outside of NIPH. Unless the Division for Health Services is going to invest in creating its own ML development – which is currently occurring in other divisions – we must maintain collaboration with important ML development actors (e.g. University College London, Epistemonikos) and other evidence synthesis institutions (e.g. the UK's National Institute for Health and Care Excellence, Sweden's Statens Beredning för Medicinsk och Sosial Utvärdering) as a form of horizon-scanning for new ML functions, applications, and best practices related to capacity-building. These collaborations are also an important pillar in internal capacity-building.

Presenting and publishing our findings (see Dissemination) and maintaining existing research collaborations are important ways to grow our current network. This is needed to continue developing our own expertise. We will seek to identify other evidence synthesis institutions who are implementing ML to explore possibilities for knowledge sharing and mutual training, e.g. particularly those working with knowledge graphs and networks, and automated data extraction. A concrete example of this is the current exchange of reviewer-oriented ML guidance material with the University of North Carolina.

Internal learning and knowledge transfer

There is also a considerable amount of internal ML expertise and activities occurring within NIPH, e.g. in the Division of Infection Control and Environmental Health, Division of Health Data and Digitalisation, and IT. These activities are not specific to ML, but can likely be transferred to and utilized by HTV. The first institutional ML and big data networking meeting was conducted on 23.6.2021, and next steps are being decided.

Dissemination

Dissemination is key to building our network, knowledge-sharing, and internal capacity-building. We plan to disseminate our work through presentations internally at NIPH and externally at conferences or presentations at collaborating institutions. When relevant we will plan on publishing our findings as journal articles.

Evaluation, testing, and validation activities

New ML functions and applications will often need to be internally evaluated, both for acceptance and for applicability to HTV workflows, and potentially modified. These evaluations will: provide the evidence base for the functions we move forward with, form the basis for recommendations for further exploration or development, and inform capacity-building activities (e.g. which functions need how-to guides in Norwegian? which can be learned through existing guidance?). Through these evaluations, we will contribute to innovative and reproducible research.

Specific evaluation activities that we recommend prioritizing:

- Further application of existing functions, e.g. identification of search terms through automatic text clustering and other unsupervised ML
- Evaluation of upcoming functions, e.g. automated data extraction
- Support for a librarian group in evaluating Microsoft Academic Graph or similar products.

ML in tandem with workflow optimization

As discussed in the team report, ML's benefits appear maximized by corresponding workflow changes. We will continue mapping and exploring areas where ML and other types of automation meet. There is a lot to learn outside of HTV, for example, NIPH's *Folkehelseprofiler* create tailored, updated reports, which could be particularly salient for the creation of both ML-team materials and review products themselves. Workflow changes can be disruptive, however, and ideally, we will cooperate with or receive guidance from other employees or teams experienced in change management, learning organizations, and so on.

Implementation and capacity-building among project groups

The ML functions evaluated as effective (and acceptable, ethical, and of equal or higher methodological quality) are the ones which the team will help implement in project teams and build capacity of these teams to use independently.

Capacity-building strategy for HTV and the Division for Health Services

First, for functions for which we have the strongest evidence base, we will scale up the use of the stand-alone training materials we have developed. We will continue to offer a team kick-off meeting (together with the EPPI superuser, when relevant) for teams while they are writing their project plans. These meetings facilitate an understanding of both software options and ML functions that could be relevant to their project and why, and feed into our needs assessment process to help organize the ML team's activities and recruit rotating members.

The first line of support will be EPPI Centre's skilled support team, with whom we have a commercial relationship and existing research collaborations. If EPPI Centre is unable to help or if there are language barriers, a ML team member will provide support. We will transition away from providing one-on-one guidance. This type of support will continue until we reach a threshold of confident/trained project leaders (for example 75% or 15 people), representing a level of "saturation" of skills; these project leaders should be confident enough to a) implement the ML functions we support with only modest assistance from our team, and b) train their project members in using them. We recommend that a ML-naïve project leader joins a project group with an experienced ML project leader to co-lead the ML aspects of the project, before this ML-naïve project leader begins implementing ML functions in their own projects. This suggestion is based on the training hand-off procedure we used to build capacity in the ML team.
Second, after we have reached a threshold of confident and trained project leaders, our focus will shift to capacity-building amongst other project members. This will be accomplished through larger training sessions and group opportunities. Some possibilities are during new employees' course, seminars, and potentially non-project-based trainings.

Specific capacity-building activities that we recommend prioritizing:

- Create a visual, guidebook, or decision tree for project leaders to use to identify when, where, and why it is appropriate to use ML in evidence synthesis products. This visual could map ML functions, review types, review phases, and review characteristics, as well as available training materials and user-friendly summaries, such that project leaders would be able to quickly see the ML functions recommended for their specific review. It could also be the landing site of an open ML resource page or Teams room for employees.
- Templates for standard sentences, reporting, and PRISMA figures for protocols and full reports, sent to the team responsible for the methods handbook.
- Suggested language for responding to peer-reviewer or commissioner questions.
- Extend implementation to qualitative evidence syntheses. Every function we have evaluated so far utilizes naturalized language processing, i.e. learning from text with

inherent meaning rather than numbers, and there is no reason to limit ML to quantitative reviews.

In addition to the technical/software competence needed to implement ML functions, employees need a basic, but fundamental, understanding of ML. Depending on ML function, employees must have some level of understanding of the mechanisms in order to understand the output, and potential pitfalls: what one can expect, what can go wrong, what can misunderstood. Some of our user guides provide this kind of knowledge, whereas other functions, such as priority screening, may be more intuitive and need less background knowledge. There are several options to provide this education:

- Re-use the internal syllabus and teaching materials from 2021.
- Recruit experts from outside of the ML team, such as those who participated in our June networking meeting from the other divisions, this network's contacts at NTNU, UiO, and OsloMet, to provide stand-alone trainings.
- Embed ML experts within the team as rotating members.
- Look for ML and programming skills in new hires.
- Start a ML and big data network that goes across NIPH.

Dissemination and communication activities

Dissemination – unidirectional information flow from the ML team to others in the Division for Health Services – is necessary to increase understanding, buy-in, and trust. We will disseminate our findings, including benefits and challenges, as well as possibilities with ML, with the aim of demystifying ML.

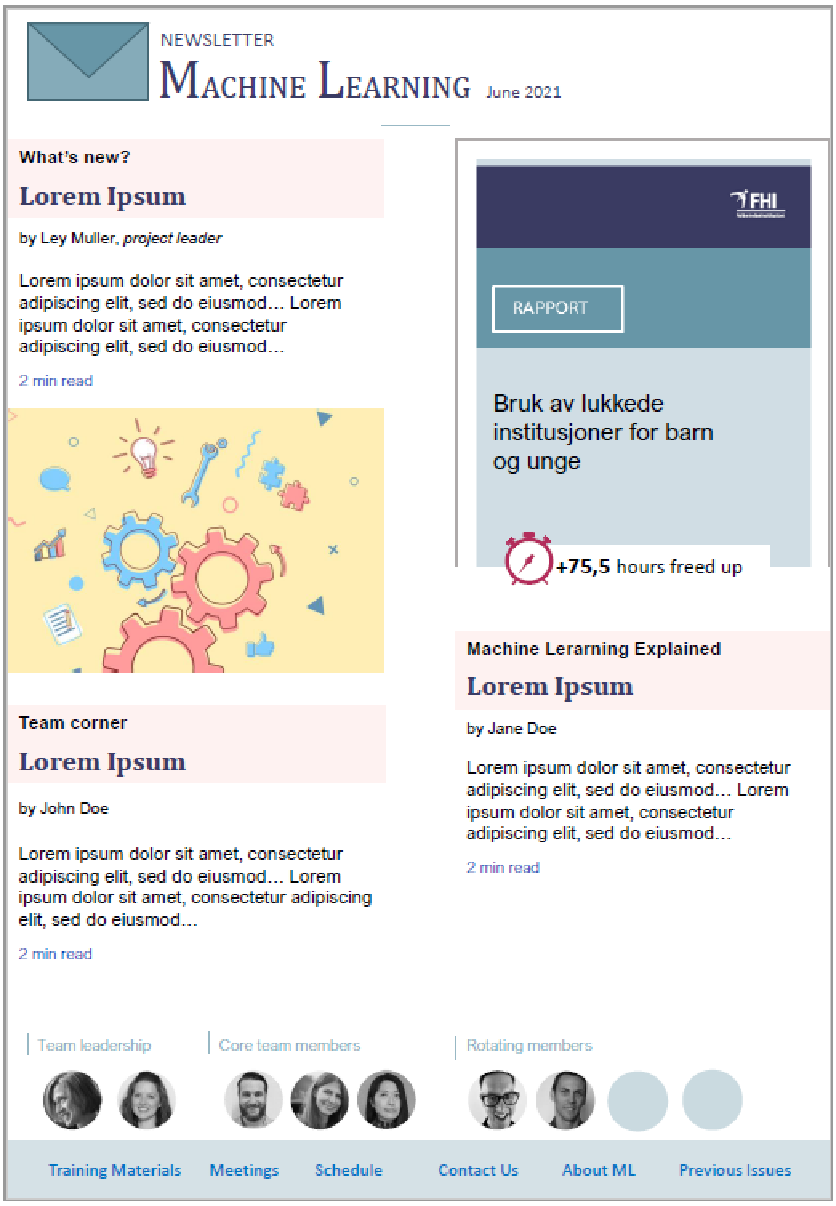
It is also important to facilitate communication, i.e. bidirectional information flow between the Division for Health Services and the ML team. The goal here is to increase engagement, and this includes hearing and engaging with critical voices, understanding Division for Health Services's and individual researcher needs, and inviting innovation and problem-solving *outside* of the ML team.

We suggest a three-pronged approach:

1. Utilize existing communication channels such as the weekly meeting in the Division for Health Services (ukestart) and bi-weekly meetings in HTV (klyngemøte) to disseminate, to ensure that we reach most of our intended audience.
2. Produce a short newsletter or other type of regular communication mechanism to disseminate. See figure 2 below for an example. The newsletter could have distinct sections: a) announcement of new, interesting, and relevant findings of our own evaluations, or from other scientific articles; b) a visual presentation of a statistic or achievement; c) a team corner in which a team member discusses a topic of their own choice, to stimulate curiosity; and d) a ML term explained, to familiarize readers with ML expressions and jargon.
3. A more dynamic method of communication may be needed, such as a regular "ML lunch", to stimulate critical engagement. We could use this communication channel to invite critical feedback and engagement, so that employees know that there is a devoted, open line of communication for their ideas of how the team could be working better. The focus here will be on listening to employee feedback, and could

serve the additional purpose of identifying interested project leaders who could become rotating team members.

Figure 2: *Example of possible newsletter to communicate ML activities*



Building expertise among ML team members

Related to capacity-building in HTV and the Division for Health Services is building expertise among ML team members. The ML team needs to build their skills relating both to ML and to implementing and scaling up innovative (and sometimes disruptive) technologies; if the 2021-2022 team is comprised of new members, they will additionally need time to build basic competence. The team's daily activities will provide hands-on practice. However, formal training maybe be more effective, and we anticipate internal training within the team and through external training or courses, conferences, cross-division collaboration, and collaboration with other institutions as opportunities. Examples of relevant training or skills that that would highly benefit the team: innovation leadership, change management, managing disruptive technologies, project management in general, and more advanced courses related to novel ML functions within evidence synthesis.

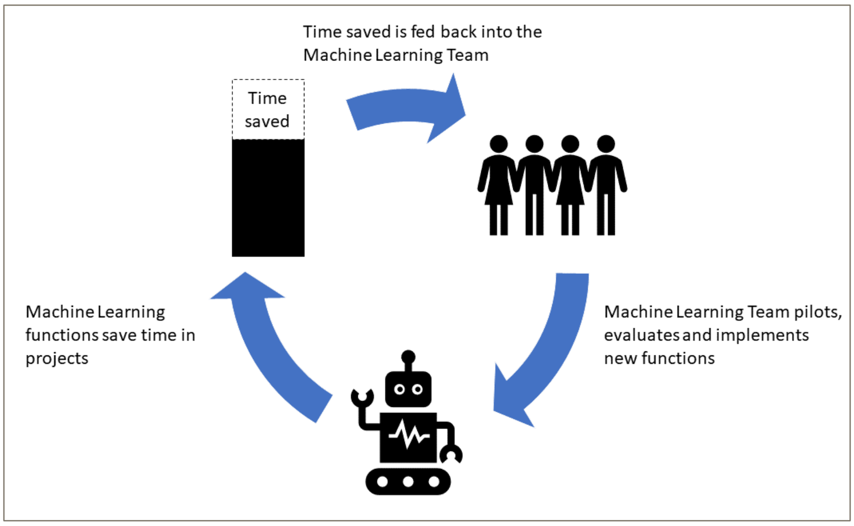
Reporting to leadership

We will give short semester updates to the leadership group on ongoing activities and results. In Q1 2022 we will deliver a status report where we can present changes to the strategy or team members as needed.

Resource and time needs

Resources for the team can be drawn from the time that the current ML team has saved HTV by utilizing ML in previous and current projects; i.e. that time saved through ML implementation is then "banked" and used by the new ML team (see figure 3 below).

Figure 3: *Illustration of resource loop*



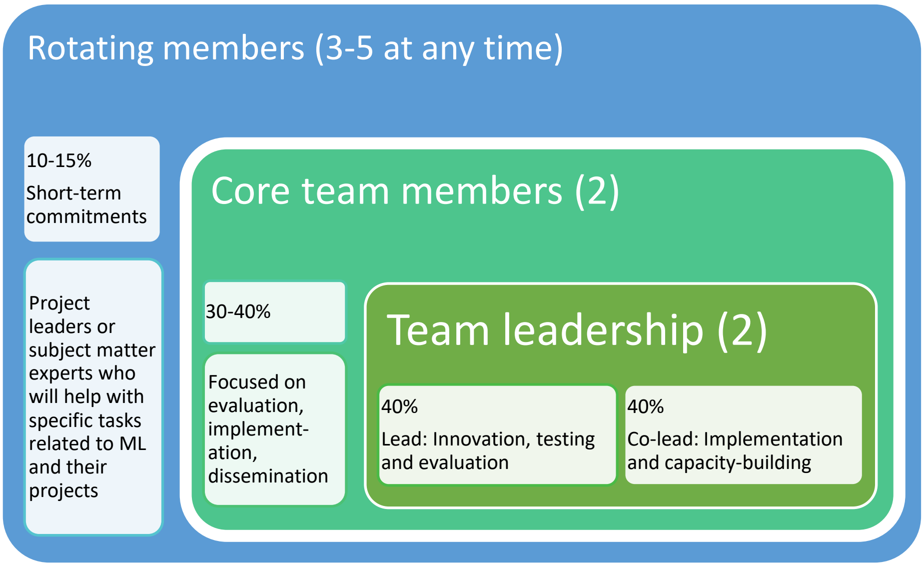
We propose that the members of the core group remain small and follow the traditional team structure. For an outline of the team structure, see figure 4.

The team leader will have responsibility for innovation, testing and evaluation, externally-focused capacity-building, and team administration. The co-lead will be tasked with implementation and capacity-building (including dissemination) and will step in as team leader when needed. As the ML team advances, implementation and capacity-building roles will be more and more distinguished from innovation and testing, and both activity groups need to be prioritized by a lead with specific skills.

We also propose an "extended team" of rotating members, such that the ML team becomes more porous. We will identify and invite project leaders or subject area specialists to be involved for shorter periods of time, and potentially with a lower time commitment. We anticipate 3-5 of these team members at any given point. Activities could include evaluating a novel ML function in their project (and being trained to collect data), advising on innovation diffusion, bringing a novel project type in (such as a QES) that integrates as many ML function as possible, or updating standard reporting language for the methods handbook. Core team members roles are then to support these distinct activities, analyze evaluation data, and disseminate further.

By including rotating members, we are also attempting to increase the number of employees who understand and champion ML functions, which will help prevent the core team from becoming a closed environment.

Figure 4: *Possible team structure for a future ML team*



Librarians have created their own working group to test ML functions relevant to their tasks, and should proceed in parallel to the ML team. The librarians will pilot the training documents created by the ML team and EPPI Centre to independently implement ML functions. The ML team will provide support and advice related to methodological decisions and overcoming roadblocks.

We also see a benefit in giving the ML team a budget to facilitate learning, dissemination, and collaboration. This would support team members participating in courses, conferences, and meetings with identified institutions working on the same objectives.

Risk assessment

We have conducted a risk assessment of six organizational, technological, and human resources risks. For risks assessed as medium or high (a combination of likelihood of the risk occurring and its impact), we have suggested mitigation strategies, and the subsequent risk after mitigation. See figure 5 and table 1 below.

Figure 5: *Illustration of risk assessment with likelihood and impact*

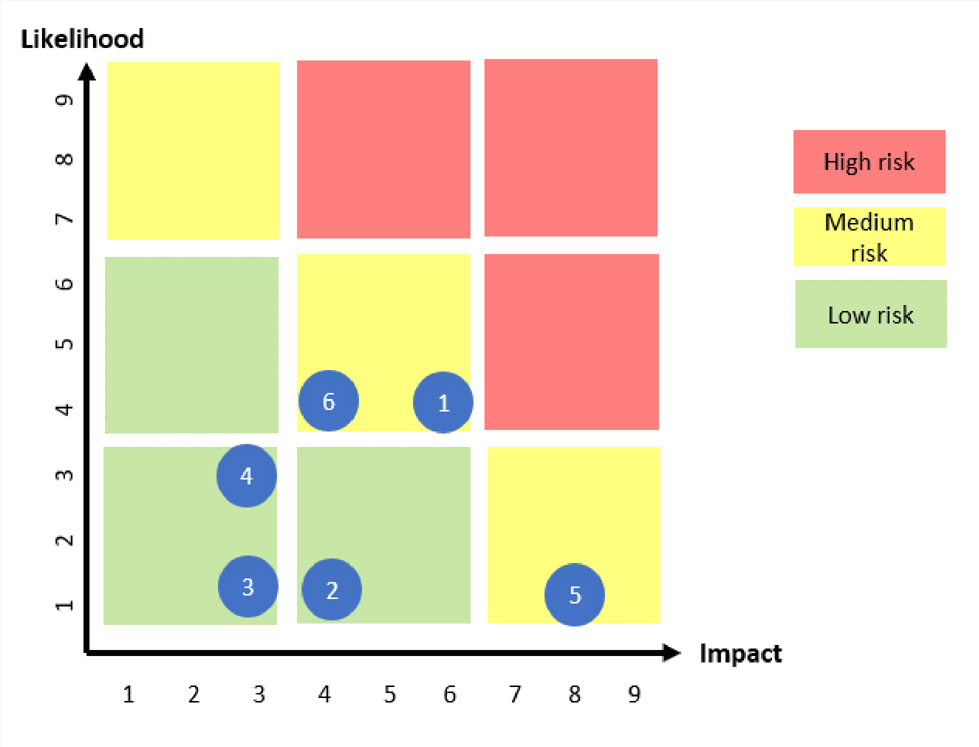


Table 1: *Description of risk assessment*

#	Description of risk	Risk level	Risk mitigation	Risk level after mitigation
1	Innovation risk: The team has inadequate time to identify new functions/applications, such that the team primarily provides training help, and over time on less effective functions.	Medium	Separate innovation activities from testing and implementation activities; establish clear responsibility for innovation activities.	Low
2	Evaluation risk: insufficient project interest to evaluate new functions.	Low	--	--
3	Implementation risk: insufficient project interest to implement effective functions.	Low	--	--
4	Capacity-building risk: Lack of interest among project members or leaders to be trained in ML functions, or to utilize them.	Low	--	--
5	Software provider (EPPI Centre) ceases to deliver necessary software (EPPI Reviewer) for the majority of our ML functions.	Medium	Explore new software, exploit existing NIPH human resources to learn new functions/software.	Low
6	Loss of project members.	Medium	Begin with sufficient team depth; establish a co-leader.	Low

Long term developments

This strategy suggestion has been created based on the situation today and is informed by the ML team's six months of experience. It should be updated in accordance with institutional changes or developments in the field.

Our priority is to implement and consolidate ML functions within HTV. However, once we establish that a ML function is effective and should be implemented, we will expand beyond HTV to other groups within NIPH producing evidence synthesis products, such as Global Health and several projects within other divisions. We expect this to happen within the next year. We envision transitioning from a cluster-level to division-level focus at this time.

The ML team is unique among other employees working with ML in NIPH, because it is closer to the application of ML (and communication with end users) than to the development. That is, we are software and ML users, rather than ML programmers. In the future, with stronger knowledge transfer from programmers in other divisions or new hires with programming capacity, we need to decide the strategic value of becoming developers of our own ML functions.

Published by the Norwegian Institute of Public Health
September 2021
P.O.B 4404 Nydalen
NO-0403 Oslo
Phone: + 47-21 07 70 00
The report can be downloaded as pdf
at www.fhi.no/en/publ/