# The implications of artificial intelligence and machine learning in health financing for achieving universal health coverage

*Findings from a rapid literature review*

World Health Organization

# The implications of artificial intelligence and machine learning in health financing for achieving universal health coverage

*Findings from a rapid literature review*

World Health Organization

The implications of artificial intelligence and machine learning in health financing for achieving universal health coverage: findings from a rapid literature review

ISBN 978-92-4-006401-0 (electronic version)
ISBN 978-92-4-006402-7 (print version)



**Suggested citation.** The implications of artificial intelligence and machine learning in health financing for achieving universal health coverage: findings from a rapid literature review. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.

**Cataloguing-in-Publication (CIP) data.** CIP data are available at http://apps.who.int/iris.

# Contents

# Acknowledgements

This paper was prepared by Maarten Oranje and Inke Mathauer. Valuable input and comments were received from Andrew Mirelmann, Fahdi Dkhimi, Natschja Ratanaprayul, Bruno Meessen (WHO Geneva), Grace Kabaniha and Jaidev Anand (WHO Country Office, India), Triin Habicht (WHO Regional Office for Europe) and Lluis Vinals Torres (WHO Regional Office for the Western Pacific).

WHO also thanks Joe Kutzin for his helpful comments and overall inspiration as well as the members of the WHO Health Financing team for their reflections during a presentation of a draft version of this paper.

Finally, WHO gratefully acknowledges financial support received from the Government of Canada.

# Executive summary

Digital health is a rapidly expanding area of interest for both research and practice, and within this area digital technologies for health financing receive increasing attention. Among many other digital technologies, digital health encompasses the use of artificial intelligence (AI) and big data. When we explore the multiple potential benefits of digital technologies for health financing, we also need to pay attention to the possible risks and challenges. While there is an abundance of publications on the application of AI for different areas in health in general, the topic of AI in health financing has received much less attention.

To contribute to closing this gap, this paper – which is based on a rapid literature review – provides an overview of the current applications of AI and machine learning (ML) for health financing functions and tasks in research, policy and practice. The different uses of AI and ML are mapped in relation to various health financing functions and tasks as well as to the objectives of universal health coverage (UHC). The main purpose is to identify the type of research and policy questions that are being studied and to explore key issues and implications in relation to progress towards UHC. For the mapping exercise, the authors searched in PubMed, Google Scholar and Google between January 2000 and December 2021. In total, 38 publications were selected and data extracted.

This rapid literature review exposed the broad scope of health financing-related subjects to which ML approaches are applied. These are grouped into six categories: prediction of health expenditure, risk scoring, claims management and fraud detection, identification of households for targeted policies, health needs informed benefit package design, and analysis of the effects of health coverage scheme design on health service utilization.

The review shows that ML is primarily applied to health insurance-related topics and, more specifically, to private (commercial) health insurance. Only one paper relates to a low-income country, 15 are based on data from middle-income countries and 22 are from high-income countries. The majority of papers (34 out of 38) constitute modelling exercises or a proof of concept. Some of these papers are also driven by a public policy orientation, discussing the benefits and risks of ML in health financing and/or suggesting policy options in the direction of UHC.

Most ML uses relate to the pooling or purchasing functions, or both. The authors did not find any publication specifically about the use of ML to raise revenue for health – such as the setting or collection of health insurance contribution payments. Two main topics were found to be dominant in this literature, namely: 1) prediction of high-cost patients and health expenditures; and 2) fraud detection in health insurance claims management. Overall, the use of ML seems to produce findings that confirm and/or are consistent with existing knowledge. The main added value of ML lies in its enhanced speed and precision or accuracy compared to traditional statistical methods.

In conclusion, the various applications identified of ML for health financing could have the potential to affect all the intermediate UHC objectives, most importantly the equitable distribution of resources (either positively or negatively) and efficiency (most probably positively). Transparency and accountability could also be improved for example through more effective detection of fraud. In addition, there may also be effects on all of the three final UHC goals: utilization in line with need, universal financial protection and quality of care. Whether those effects will be positive or negative will mainly depend on how the technology is applied. Although ML techniques can change traditional ways of working, they do not seem to alter fundamentally the thinking about health financing – or at least not yet.

For future research, topic-focused scoping literature reviews as well as UHC policy-focused country studies are proposed in order to widen the gathering and synthesis of evidence on the possible added value of AI and ML for health financing. A more cross-cutting research topic relates to the design, implementation, enforcement and impacts of regulation of AI and ML, particularly in relation to data sources and algorithm design to avoid algorithmic bias.

Ultimately, research and practice related to AI and ML for health financing must be based on and driven by a clear purpose in the public interest and by benefits to the health system, rather than particular interests of individual stakeholders and insurers. When the ethical issues related to AI and ML in health financing are addressed and feasible regulatory options are in place, there can be accelerated progress towards UHC.

# 1. Introduction

Digital health is a rapidly expanding area of interest for both research and practice, and within that area digital technologies for health financing receive increasing attention. Very little robust evidence exists on the impact of digital technologies on health financing objectives. The number of uses and studies is gradually growing but more rigorous research in this area is urgently needed (WHO, 2021b; Hanson et al., 2022). When exploring the multiple potential benefits of digital technologies for health financing, there is also a need to pay attention to the possible risks and challenges. The application of digital technologies can, for instance, negatively affect existing health financing arrangements that are conducive to universal health coverage (UHC), either by inadequate design or as a result of suboptimal implementation (WHO, 2021b).

Among many other digital technologies, digital health includes the use of artificial intelligence (AI) and big data analytics for health (WHO, 2021a). Concerns about potential risks are particularly relevant in relation to the application of AI and of machine learning (ML), the latter being a prominent type of AI. Although there are many publications on the application of AI for various areas in health, including from WHO and other international organizations (OECD, 2019; UNESCO, 2021; WHO, 2021c), the topic of AI in health financing is typically mentioned only in a cursory way.

To help close this gap, this paper – which is based on a rapid literature review – provides an overview of the current applications of AI and ML for health financing functions and tasks in research, policy and practice. There is a focus on ML because this is currently found to be the dominant approach by which AI is used in the health financing domain. The different uses of AI and ML are mapped in relation to various health financing functions and tasks as well as to UHC objectives. The main purpose is to identify the type of research and policy questions that are being studied and to explore key issues and implications in relation to progress towards UHC. The paper assesses neither the types or effectiveness of ML approaches nor the quality of the evidence.

For this rapid literature review, published and grey literature as well as publications from international organizations in English were screened. Searches were carried out in PubMed from January 2000 to December 2021, Google Scholar and Google. In PubMed, a title and abstract search was completed. Annex 1 presents the search terms and strategy for PubMed. The authors then further sieved through the papers and included those with a definite case of ML use relating to health financing functions or subtasks. Papers on health service utilization were included when this was explored in relation to the design of a health coverage scheme. However, papers related to purchasing that focused on one specific disease or health intervention only (e.g. prediction of costs for patients with diabetes) were excluded from the review. Both authors independently reviewed the papers for inclusion and exclusion; differences were resolved through consensus. In total, 38 publications were selected and data extracted.

# 2. Artificial intelligence and machine learning

AI can be described as "the ability of algorithms encoded in technology to learn from data so that they can perform automated tasks without every step in the process having to be programmed explicitly by a human" or, in a shorter version, "the performance by computer programs of tasks that are commonly associated with intelligent beings" (WHO, 2021c). All forms of AI build on the availability of large quantities of data and the capacity of algorithms to analyse those data rapidly and translate them into information, conclusions and actions.

ML is a type of AI and is based on the use of statistical and mathematical modelling techniques to define and analyse data (WHO, 2021c): algorithms are trained to recognize patterns or to make estimations or predictions without human interference. The four main categories of ML are supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning (Panesar & Panesar, 2020) (see Box 1).

### Box 1. Categories of machine learning

For **supervised learning**, the data used to train a model are labelled (the input and respective outcome variables are known) and the model derives a function from the training data that can then predict outputs from new input data. Supervised learning problems can be further grouped into regression problems (where the output variable is a continuous value, such as "future health expenditure"), classification problems (where the output variable is a distinct category, such as "a person with high health risk" or "a person with low health risk") and forecasting problems (where predictions are made on the basis of past and present data). An "ensemble" is a type of supervised learning that combines multiple different ML models to predict an outcome for a new sample (Panesar & Panesar, 2020). Supervised learning techniques include linear regression, logistic regression, Bayesian networks (such as Naïve Bayes), random forest, k-nearest neighbor (k-NN) and support vector machines.

**Box 1. (contd.)**

**Unsupervised learning** does not use labelled data but involves the identification of hidden patterns in the data by a machine (WHO, 2021c). There are two main types of unsupervised learning problems. First, for clustering problems, elements (e.g. patients or people in a population) are grouped in such a way that elements within the same cluster are more similar to each other than to elements from another cluster (e.g. a cluster of people likely to need financial support versus another cluster with different needs). Second, in association or correlation problems, the probability of the co-occurrence of items in a data set is predicted (Panesar & Panesar, 2020) – for instance the likelihood that a person with certain socioeconomic features will not be able to pay user charges. Unsupervised learning techniques can also be used to detect anomalies such as suspicious claims. Some of the common unsupervised learning techniques are hierarchical clustering (clusters are ordered hierarchically) and K-means clustering (similar data points are grouped together in a set number of clusters [k] discovering underlying patterns) (Usama et al., 2019).

**Semi-supervised learning** combines elements of both supervised and unsupervised learning: a small amount of labelled ("training") data is combined with a large amount of unlabelled data. This is also referred to as a form of weak supervision.

**In reinforcement learning**, optimal actions (or an optimal policy) are learned through trial and error as well as feedback allowing the algorithm to optimize behaviour. This differs from supervised learning in the sense that correct examples are never presented and suboptimal decisions are not explicitly corrected (Panesar & Panesar, 2020).

Neural networks are series of algorithms that make up ML models, and a neural network consisting of several layers is referred to as a deep learning algorithm (Kriegeskorte & Golan, 2019).

ML requires large amounts of (good quality) data in order to produce results that are both tangible and reliable. In some cases, such data sets may qualify as "big data", which is usually defined by the three dimensions of volume, velocity and variety (Panesar & Panesar, 2020). These are complex data sets that are rapidly collected in quantities requiring unconventional storage space and methods of analysis (WHO, 2021c). In the use of data analysis techniques, there are large similarities between machine learning and data mining. Data mining is by definition a human-led process, in which the machine itself does not learn and thus it should strictly not be considered to be a form of AI. In practice, however, there is not always a strict boundary but rather a large overlap. The terms are often used interchangeably.

# 3. Findings

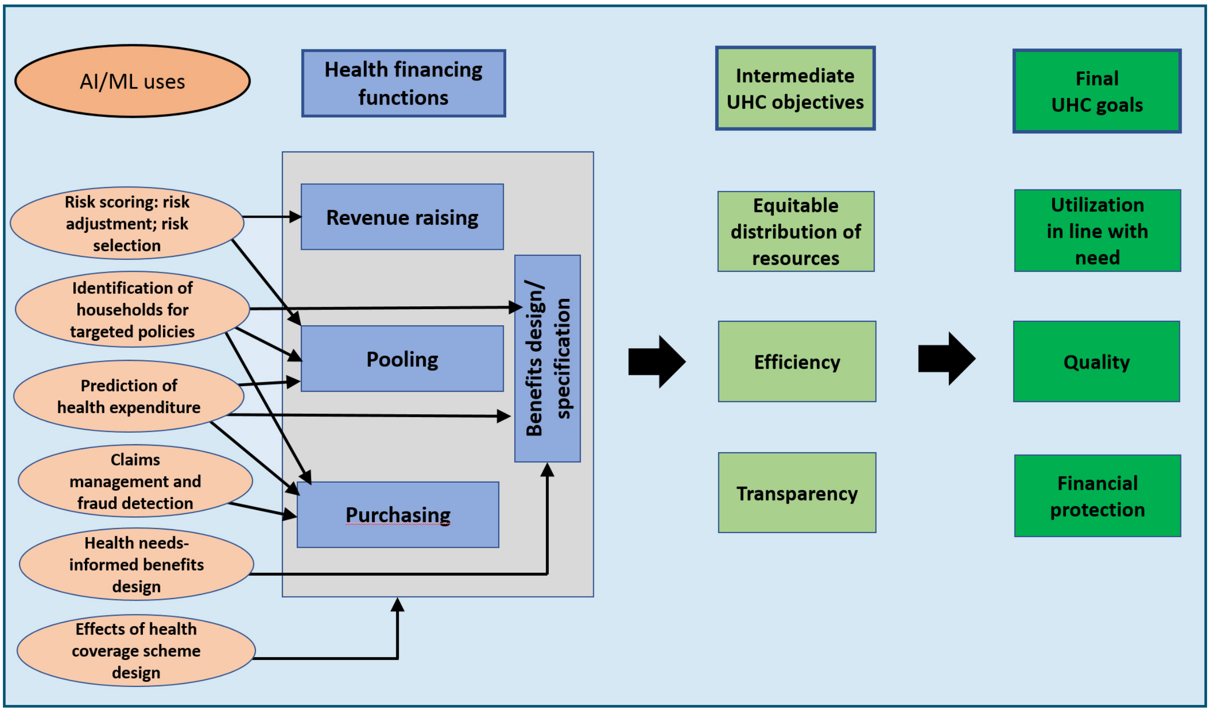
The initial intention was to group the included papers in line with the three health financing functions of revenue-raising, pooling and purchasing. However, the authors realized that there was a need to categorize the papers into more detailed subtopics, some of which relate to more than one of the health financing functions. Thus the papers were grouped into six categories related to different health financing tasks. It is acknowledged that the boundaries are not always clear-cut and alternative categorizations would have been possible. Table 1 presents the subtopics that were identified and the number of papers grouped in each category. Annex 3 summarizes all papers, the type of data used, the type of ML applied and the main findings.

**Table 1. Overview of papers on use of AI/ML grouped by subtopics related to health financing tasks**

| Topics | No. of papers |
|---|---|
| Prediction of health expenditure | 9 |
| Risk scoring: risk adjustment versus risk selection | 5 |
| Claims management and fraud detection | 13 |
| Identification of households for targeted policies | 6 |
| Health needs-informed benefit package design | 3 |
| Analysis of the effects of health coverage scheme design on health service utilization | 2 |
| **TOTAL** | 38 |

Figure 1 outlines the health financing-related topics that are supported by ML and how they link to health financing functions and UHC objectives.

**Figure 1. Mapping of ML uses for health financing functions**




## Prediction of health expenditure

The use of ML models to predict high-cost patients is a frequent research topic, with eight papers found that discuss the development and testing of specific models. The underlying purpose of these predictive models is not uniform: three papers aim to inform and improve risk adjustment approaches among insurers and, in doing so, could contribute to a more equitable distribution of resources.

Along this line, using claims data from United States health insurance plans, Kan et al. (2019) tested different regression algorithms to predict future health-care costs in older adults. They suggest that these improvements in prediction may serve to improve risk adjustment and population health management in order to address the health needs of specific population groups. A paper by Rose (2018) evaluates the average incremental annual cost related to 26 of the most prevalent medical conditions and concludes that existing approaches may underestimate the spending contributions of specific medical conditions, such as major depressive and bipolar disorders and chronic hepatitis. This is of relevance because insurers may have no incentives to invest in the prevention of diseases "if current risk adjustment methods are not capturing the true incremental effect of medical conditions […]" (Rose, 2018). Similarly, Vimont, Leleu & Durand-Zaleski (2021), using a representative sample of all medical claims data for France in 2015 and comparing different ML models to predict the health-care costs of individuals, conclude that their results can help develop better risk adjustment models (under competitive health insurance) or realize more efficient and equitable allocation of resources (under non-competitive insurance).

Another underlying purpose of such prediction analysis in the other five papers is to guide prevention activities and to target patients with particular health needs in order to reduce health expenditure, enhance efficiency in resource allocation and/or to align utilization with health needs.

Along this direction, using data from the Medicaid programme in the state of Texas, United States of America (USA), Yang et al. (2018) tested four predictive models to forecast expenditures – especially for "high-cost, high-need" patients – and showed significant temporal correlation in health-care expenditures. Authors suggest that these insights can help to direct precise preventive care to high-cost patients in order to reduce overall health-care costs and to provide care more efficiently. Similarly, Nomura et al. (2021), using public health insurance data from Japan, compared various ML models to predict the annual medical cost, with a neural network performing best. They discovered that medical cost proportionally depends on the medical cost of the previous year. This information can be used by health policy-makers to prioritize prevention activities and thereby reduce future health-care costs.

In order to inform future trials that test the effectiveness of interventions to reduce health-care utilization, Ng et al. (2020) applied ML algorithms to claims and clinical data of inpatients and outpatients of a university hospital in Singapore to predict which high utilizers will be persistent high utilizers. In a similar manner, Joedicke et al. (2019) on the basis of Swiss health insurance claims data, tested several supervised ML techniques to predict the development of patients' health-care costs in the subsequent year and to identify contributing factors, with a focus on the contribution of pharmacotherapy. The researchers stated that the findings could assist policy-makers to improve resource allocation planning by optimizing health-care services (disease management programmes) for patients with high costs or, for instance, by reducing the deductible of low-cost patients. Finally, using claims data from a private health insurance company in Australia, Xie et al. (2015) tested whether the way in which clinical codes are hierarchically structured can be utilized to improve their power to predict a patient's hospital days in subsequent years. They found that different hierarchies had no significant impact on the predictive power of the models.

Beyond the individual modelling studies, Nichol et al. (2021) conducted a systematic database search of relevant business news and academic research, identifying and categorizing ML-based predictive analytics products that claim to improve the efficiency of health care. Out of a total of 106 "products", they found 38 that predicted the cost and/or utilization of health care; these were chiefly products that predict high (cost) utilizers or that support financial risk stratification of patient populations. The authors also pointed to specific ethical and regulatory challenges arising from the use of ML to improve health-care efficiency, because reducing costs may have negative effects on quality and biases in the data may exacerbate discrimination.

All in all, the papers in this section suggest that the use of ML may generate additional insights compared to traditional statistical methods – e.g. by identifying specific patient attributes or other determinants which enable more accurate estimates of the health expenditure of groups or individuals (Rose, 2018; Vimont, Leleu & Durand-Zaleski, 2021). Importantly, the application of ML analytics to financial and claims data is not limited to health financing: it could also inform actions and improvements in service delivery. However, at the same time, such information could be used to exclude people in need or to increase their insurance contributions, leading to further fragmentation or reduced equity in resource distribution (and consequently harming UHC objectives).

## Risk scoring: risk adjustment versus risk selection

Several publications were found on ML-based risk adjustment models, all using data from the United States. Rose (2016) applied several ML techniques to improve the risk adjustment mechanism among private health insurance schemes, finding that an ensemble (using a weighted average of multiple algorithms) outperformed classical regression and other individual algorithms. Rose, Bergquist & Layton (2017) focused on one specific component: the prescription drug formulary. They demonstrated that a small number of variables can be used to identify "unprofitable" enrollees and therefore called for regulation to mitigate this risk. Similarly, McGuire, Zink & Rose (2021) challenged the usual method of improving risk adjustment formulas by increasing the number of variables: with a reduced number of attributes, they claimed performance that is equal to or better than the existing risk adjustment model.

The rationale of the papers referred to above is to improve equitable resource distribution across multiple (competing) pools. However, at the same time, risk scoring could also be used to optimize risk-rated premiums and risk selection strategies (of voluntary health insurance schemes), which could increase inequities in financing and financial protection. Two papers were identified that hint at risk scoring for risk selection purposes. Hileman & Steele (2016) used voluntary health insurance claims data from the USA and tested a random forest model, which had been trained with Medicare data, on its risk-scoring performance. They concluded that that this ML-based risk-scoring model may also have potential for voluntary health insurance. Although not explicitly presented as risk scoring, the approach of Discovery, one of South Africa's largest private health insurance companies, as summarized by Mrazek & O'Neill (2020), also points in that direction. Discovery applies ML to members' data received from business partners on their supermarket purchases, activities in fitness firms and health-care utilization, in order to assess the level of customers' healthy behaviour. As a "shared value" business model, members are offered financial rewards (i.e. premium paybacks) for healthy behaviour, while the insurer reduces the average risk of its member pool when members behave in a healthier manner. From a health-financing perspective, however, such models have a real risk of desolidarization – which would discriminate against those in greatest need of health coverage.

## Claims management and fraud detection

Claims management and fraud detection supported by ML techniques is an area that has received a great deal of attention from various academic disciplines such as (health) economics, public administration, computer science, technology and (information systems) engineering.

The authors identified 12 papers that assess specific ML approaches to claims management and fraud detection in health insurance. While 10 papers focus on the identification of fraudulent claims submitted by providers (see below), two papers focus on earlier stages of the claims management process. Using hospital data from China, Liu et al. (2021b) tested five common classification algorithms for designing and updating data-based grouping of diagnosis-related groups (DRGs), with the aim of reducing the practice of upcoding that occurs in expert-based grouping. The researchers showed that the classification performance of data-based grouping is similar to that of expert-oriented grouping, depending on the choice of suitable algorithms, and shows a higher level of transparency with lower costs in designing and updating the grouping. This claim of high transparency (i.e. no dependence on expert judgement) can of course be challenged, as the use of algorithms comes with its own risk of bias and lack of transparency. The inner workings of an algorithm and the underlying choices are not always easily explained to non-experts (Panch, Mattie & Atun, 2019). Araújo, Santana & de Santos Neto (2016) demonstrated the potential of an ensemble (combining three classification techniques) to support the process of pre-authorization of dental claims by professional reviewers in Brazil, so that the less complex cases could be evaluated fully automatically.

The other 10 papers focus on the identification of fraudulent claims submitted by providers. Most of these papers use data from voluntary health insurance: Lu & Boritz (2005) on Canada; Kirlidog & Asuk (2012) and Kose, Gokturk & Kilic (2015) on Türkiye; Kumar, Ghani & Mei (2010) and Ekin et al. (2013) on the USA; Sun et al. (2020) on China; and Ortega, Figueroa & Ru (2006) on Chile (for which data from mandatory private health insurance are used). Three papers use data from public health insurance, namely: Shin et al. (2012) on the Republic of Korea; Sowah et al. (2019) on Ghana; and Joudaki et al. (2016) on the Islamic Republic of Iran. Of the 10 studies, seven papers were found to apply supervised or semi-supervised learning models, while three papers used unsupervised methods. Ekin et al. (2013) did not focus on fraud by providers exclusively, but used an approach referred to as co-clustering to zoom in on suspect combinations of providers and beneficiaries.

All the papers demonstrate various advantages of the use of ML for claims management and fraud detection – such as automated detection, improved accuracy of classification into legitimate and fraudulent claims, early detection of abnormal or problematic claims that require specific attention, a higher precision and sensitivity than existing approaches, and a higher detection rate. Automated detection reduces the time needed for claims processing and enables the subsequent investigation process and the work associated with reprocessing claims to be minimized, thus avoiding payment errors and reducing overall administrative costs. Above all, it is argued that ML-supported claims management and fraud detection can help increase efficiency and bring down health-care costs.

Most papers do not systematically compare the performance of ML models with traditional approaches without ML but focus instead on a performance comparison between different algorithms. However, some papers attempt to quantify the potential cost savings from the use of ML for claims management or fraud detection; Ortega, Figueroa & Ru (2006) reported cost savings of some 10% in Chile. Kumar et al. (2010) asserted potential savings for a typical United States health insurer of US$ 15–25 million per year. While these findings may not be generalizable, they do give an indication of the range of approximate cost reductions. In addition to enhancing efficiency, effective fraud detection may also contribute to transparency and accountability. Nevertheless, the use of algorithms could also possibly compromise transparency on how results are generated.

One paper related to fraud detection outside the realm of health insurance, focusing on health facilities in Zambia receiving performance-based payments. In this paper, Grover, Bauhoff & Friedman (2019) compared the performance of different ML techniques in identifying health centres that would need to be targeted for verification in order to reduce fraudulent claims and unnecessary verifications.

## Identification of households for targeted policies

Six papers were found that discussed ML based approaches to identify and target households in need of health coverage or other social assistance schemes. Improved targeting would serve the dual purpose of enhancing financial protection and improving efficiency. Applying K-means clustering, Mumtaz & Whiteford (2021) used socioeconomic survey data from Pakistan to cluster households and geographical locations according to their need for government assistance. They found four clusters of households that exhibited common patterns related to their risk exposure in terms of loss of income and property, unemployment, disaster and disease. Cinaroglu (2020) compared the performance of different ML models to predict which households in Türkiye will face out-of-pocket expenditure (OOPE). A similar question – namely which households are at risk of catastrophic expenditure as a result of OOPE in long-term care in Spain – was studied by García-Centeno, Mínguez-Salido & del Pozo-Rubio (2021), who compared a range of ML classification methods for prediction, ascertaining that the more complex models outperform the simpler ones.

Using data from the Rwanda Integrated Living Conditions Surveys, Muremyi et al. (2020) compared the performance of five different ML models to predict household OOPE and found that total household consumption is the best predictor variable for OOPE. While their main conclusion is to rightly argue for more public spending, at the same time, these insights may also help to tailor cost-sharing rules for the poor. Davis et al. (2021) used patient billing records from the USA to test different ML models to predict the likelihood of patients not paying the OOP balance of their bill for their emergency department visit, with one model correctly predicting 87% of the unsuccessful payments. It was argued that this could improve the targeting of beneficiaries for government assistance schemes, minimize patient frustration and support the billing process. Similarly, Donnenberg, Hernandez & Normolle (2021) applied supervised learning to predict which patients are unable to afford prescription medication, using national survey data from the USA. Eight key factors were identified that are associated with cost-based prescription refusal – most importantly low income and chronic medical conditions. The authors pointed out the need to increase prescribers' awareness of the scale of this refusal – and the reasons for it – in order to increase the efficacy of pharmacotherapy and lessen the patient's cost burden. The latter point, taking it further, may also point to differentiated cost-sharing policies.

In summary, the papers in this section point to the potential for using ML algorithms for the identification and targeting of households in need for assistance. The underlying rationale for discerning and describing the profiles of households is to inform financial protection policies, cognizant of the final UHC goal of universal financial protection. These studies reveal important insights as to who faces vulnerability, for what reasons, and which ML techniques have greater predictive power. The studies do not, however, fundamentally change the nature of research questions or challenge the existing body of evidence: previous knowledge is mostly confirmed.

## Health needs-informed benefit package design

Three papers were identified that sought to tailor or to increase benefits, including by refining access conditions according to people's health needs. Using insurance and medical records of hospital employees with group health insurance in Pakistan, Matloob et al. (2021) tested a K-means clustering methodology with the purpose of informing the design of health needs-based benefit packages for individual hospital employees, as opposed to the current practice of providing benefits based on an employee's position in the hospital hierarchy. On average, the ML-based approach resulted in a 25% increase in benefit amounts per employee. Dos Santos, Dias & Filho (2021) applied unsupervised learning (clustering) to nationwide health survey data of Portugal in order to identify and describe segments within the uninsured population. The researchers suggest that the specific characteristics of the health care needs of these different segments could be used to inform the design of benefits and public policies to improve access to health care for specific population groups. The papers attest to the potential of ML to contribute to bringing health-care utilization more in line with need. In doing so, the intermediate objective of efficiency may also be served. Kasy (2018) re-used data from the RAND experiment and applied a supervised learning model to determine optimum tax rates and co-insurance rates (i.e. cost-sharing determined as a percentage of the total bill), finding a much lower optimum co-insurance rate with regard to the effects on the level of health-care expenditure compared to previous studies using the same data and traditional statistical approaches.

## Analysis of effects of health coverage scheme design on health service utilization or health status

Two papers on the use of ML to assess health service utilization and health status in relation to the type of health coverage were reviewed. Kreif et al. (2020) apply supervised learning (causal forests) to explore treatment-effect heterogeneity. This group of researchers evaluated the effect of mothers' type of health insurance enrolment (subsidized, contributory versus no insurance) in Indonesia on the rate of skilled birth attendance and infant mortality. This information could help optimize programme design (e.g. targeted policies for different population groups). Similarly, in order to explore the heterogeneity of effects, Chen et al. (2021), equally using causal forests (and propensity score-matching) on data from a national survey in China, studied the effects of the Urban and Rural Resident Basic Medical Insurance (URRBMI) on health status. If the health status or service utilization consequences of the design of the health coverage scheme are better understood, as these studies aspire to achieve, this may lead to a more equitable distribution of resources and utilization in line with need. However, it is not yet clear from the current evidence whether ML-based models produce more reliable effect estimations than conventional methods.

# 4. Discussion

This rapid literature review exposed the broad scope of health financing-related subjects to which ML approaches are applied and which have been grouped into six topic categories. The review shows that ML is primarily applied on health insurance-related topics – and more specifically with regard to private (commercial) health insurance. Among the 30 papers that cover health insurance-related topics, 15 focus on voluntary private health insurance and four others on compulsory private health insurance, while 11 papers are in one way or another related to public health insurance or public health coverage schemes.

The authors noted a wide geographical range to which ML studies are applied: while 12 papers are specifically based on data sources from the USA, most other regions are also covered – including Asia (14 studies, using data from nine different countries), Europe (four studies), Africa (four studies), South America (two studies), North America (one study from outside the USA) and Australia (one study). Only one paper relates to a low-income country, 15 are based on data from middle-income countries and 22 are from high-income countries. The majority of papers (34 out of 38) constituted modelling exercises or a proof of concept in a large variety of journals ranging from health care and medicine to (health) economics and policy, health informatics, computing and (computer) engineering. Some of these papers also had a public policy orientation, discussing the benefits and risks of ML in health financing and/or suggesting policy options in the direction of UHC. Two papers focused entirely on the analysis of existing policy. It is interesting to note that eight papers were published between 2000 and 2015, another eight papers between 2016 and 2018, and the remaining 22 papers between 2019 and 2021, demonstrating the exponentially growing interest.

This rapid review of the literature did not reveal any evidence of entirely new or different research questions on health financing, unusual perspectives or novel findings. In most instances, the use of ML seems to produce findings that confirm and/or are consistent with existing knowledge. The main added value of ML lies in its enhanced speed and precision or accuracy compared with traditional statistical methods, also due to the fact that ML can be more easily applied to large volumes of data.

Most uses of ML relate to the pooling or purchasing function, or both. The authors found no publication explicitly on the use of ML for raising revenue for health. However, the search did not go beyond health financing, and one assumes that there are more papers on general revenue-raising and optimal taxation policy (e.g. Kasy, 2018) or on beneficiary targeting for tax rebate schemes to reduce inclusion errors (e.g. Andini et al., 2018).

Two main topics were identified as dominant in this literature, namely: 1) prediction of high-cost patients and health expenditures; and 2) fraud detection in health insurance claims management. Regarding the first topic, this review suggests that ML approaches provide more accuracy than traditional statistical methods, while also revealing patterns in the available data that could otherwise have remained hidden. Identifying specific patient attributes or other determinants may enable us to have more accurate expenditure estimates of people with specific health needs (Rose, 2018; Vimont, Leleu & Durand-Zaleski, 2021). This could be beneficial from a system perspective if used for better costing or for tailoring benefits as well as service delivery modes, including enhanced access conditions for disadvantaged and vulnerable population groups. At the same time, such information could also easily be used to exclude persons in need or to increase their insurance contributions, leading to further fragmentation or reduced equity in resource distribution (Ho, Ali & Caals, 2020; Thesmar et al., 2019). The focus on voluntary health insurance in the papers suggests that the risk of solidarity being undermined is real.

With regard to the second main topic – claims management and fraud detection – the advantages of ML seem more unequivocal. The availability of digitalized claims with granular information on provider activities is a prerequisite for the use of ML techniques and automated processing and analysis of claims. This can accelerate the claims management process and result in improved identification of outliers and classification of potentially erroneous or fraudulent claims. This is also reflected in various overview and review papers (Hassan & Abraham, 2013; Dua & Bais, 2014; Waghade & Karandikar, 2018; Thesmar et al., 2019) that we found as part of this growing body of literature. For instance, Thesmar et al. (2019) discussed how AI and ML can improve claim adjudication (i.e. the process of checking validity and eligibility) and fraud monitoring to detect intricate patterns within and between complex data sets.

If fraudulent claims are detected at lower costs with higher precision, and at an earlier stage, this may also lead to cost savings. Coupled with sanctions on fraudulent providers (such as reduced payments, or a loss of contract), providers may overall be incentivized to prepare better claims, ultimately leading to a decline in the rate of erroneous or fraudulent claims. On the other hand, it is important to ensure that such enhanced surveillance opportunities by purchasers or steward actors do not create a culture of fault-finding and mistrust among providers. ML-enhanced claims analysis may also help to improve decision-making on setting and adjusting payment methods and rates. Most importantly, ML-based claims analysis can enable data-driven quality improvement activities by comparing treatment practices across providers and by giving feedback to each of them, with Estonia and the Republic of Korea already operating advanced models of this (WHO, forthcoming; OECD, 2022).

The findings from this rapid literature review and our reflections above resonate with the general WHO guidance document on Ethics and governance for artificial intelligence for health. While the use of ML can help to eliminate redundant and repetitive tasks, saving time and resources, there is also concern that "the digital welfare state" could undermine access to social services and welfare, especially affecting poor and marginalized populations as new conditionalities are created (WHO, 2021c). More specifically, the use of ML in health insurance may raise ethical concerns with respect to algorithmic decision-making (WHO, 2021c). However, only a few papers in the rapid literature review discussed ethical issues of the uses of ML.

Focusing on health insurance, Ho, Ali & Caals (2020) discussed what a robust ethical and regulatory environment for big data analytics, applied in relation to health insurance, may look like and described possible safeguards. These include a clear and effective data governance framework (consistent with a human-centred approach, i.e. an approach that is based on explicit and informed consent of people, who are also provided simple ways to challenge 'automated' decisions), an accountable process to explain what and how information can be used, and the active involvement of people in the way their personal data are being managed. In other words, as is true for other sectors, data governance in health financing must ensure privacy and data protection (OECD, 2020). To guide AI and ML implementation for health financing, clear and detailed regulation is indispensable to avoid algorithmic bias and desolidarization that could lead to reduced and inequitable financial protection (Ho, Ali & Caals, 2020; Panch, Mattie & Atun, 2019). The source and nature of the data must therefore be regulated to ensure that representative and unbiased data of good quality are being used by ML models (Ashrafian & Darzi, 2018).

The main limitation of this paper is that it is based only on a rapid literature review which did not include systematic citation searching. A more focused search strategy with more sensitive search terms (going beyond the generic terms of AI and ML as well as using more specific health financing-related terms) could potentially identify a larger number of published articles and discover more grey literature as well as country case studies to the extent that these are documented. In addition, there are papers on topics which are adjacent to health financing for UHC, such as costing for hospital management, which have not been included here. For instance, the authors identified some papers on the application of ML to improve the grouping of DRGs to achieve a more efficient allocation of hospital resources (Islam et al., 2021; Liu et al., 2021a; Gartner et al., 2015). In addition there are probably many ongoing applications of AI and ML within the private sector which are not documented in scientific papers or are otherwise not disclosed for commercial reasons.

# 5. Conclusion

This review provides an overview of the application of ML to health financing. Several areas of ML use have clear benefits, such as for claims management and fraud detection where improved identification of outliers and the classification of (potentially) fraudulent claims lead to savings in both time and costs. There are also potential benefits in the use of ML for more accurate identification and prediction of high-cost patients, if applied to the tailoring of policies for vulnerable population groups. Likewise, this information could be used to tailor service delivery for these groups. Such ML-generated information could, however, also be used against the interests and needs of these population groups and could result in their exclusion from health coverage or in increased health payments, ultimately leading to reduced redistribution. If ML enhances risk selection and cost prediction in private commercial insurance schemes, it may contribute to further fragmentation, reduced or inequitable financial protection and inequitable financing.

In conclusion, the various applications of ML for health financing that were identified could have the potential to affect all the intermediate UHC objectives – most importantly the equitable distribution of resources (either positively or negatively) and efficiency (most probably positively). Transparency and accountability could also be improved through more effective detection of fraud. In addition, there may be effects on all three final UHC goals: utilization in line with need, universal financial protection and even quality, when data are used to provide feedback to health care providers regarding the quality of their services. Whether those effects will be positive or negative will mainly depend on how the technology is applied – whether for risk selection or risk adjustment, or for fragmentation or defragmentation. So far, possible ML-related improvements in health financing to accelerate progress to UHC seem mostly incremental. Although ML techniques can change traditional ways of working (e.g. in claims management and fraud detection) they do not seem to alter conceptual thinking about health financing fundamentally – or at least not yet.

Because most AI-based health applications are developed and implemented in high-income countries, their use in the context of low- and middle-income countries is more recent; therefore more robust local evaluations are needed to guide decision-making in low-resource settings (Alami et al., 2020). For future research, topic-focused scoping literature reviews as well as UHC policy-focused country studies are proposed to widen evidence-gathering and synthesis on the possible added value of AI and ML for health financing. This could include the following issues.

First, one could further explore the prediction of health expenditure for specific groups or individuals and how this may help to improve the risk equalization and risk adjustment models that are in use in various countries with multiple and/or competing health insurance schemes in order to compensate for the unequal health risk profiles of insurance members. ML-based risk adjustment models may also be used to help develop and operate capitation-based payment models.

Second, resource allocation for health from the central level to decentralized levels as part of intergovernmental transfers (such as block grants and/or conditional grants) could be informed by ML-based models in order to take account of and inform complex, parallel and inter-dependent decision-making processes and the multiple allocation formulas in use. Such models would need to be based not only on service utilization data, but also on data reflecting the real needs of people currently not using services, in order to avoid building in existing bias due to geographic inequities in service utilization.

Third, AI and ML uses and their benefits for targeting and identification of certain beneficiary groups (such as poor and vulnerable populations) for health coverage schemes could be further explored, with the idea of reducing both inclusion and exclusion errors – as, for instance, applied in India (NHA, 2019; WHO 2022).

Finally, a more cross-cutting research topic relates to the design, implementation, enforcement and impacts of regulation of AI and ML, particularly in relation to data sources and algorithm design to avoid algorithmic bias.

Ultimately, research and practice related to AI and ML for health financing must be based on – and driven by – a clear purpose in the public interest and system benefits, rather than specific interests and gains of individual stakeholders and insurers. Such UHC-conducive research and practice must involve health-financing policy experts in the design and execution of studies to, inter alia, help identify appropriate data sources and formulate policy-relevant research questions. The use of big data sets and ML is by nature expected to facilitate hypothesis generation: as algorithms can analyse large data volumes quickly and may find associations which a human researcher was not even looking for, ML may help to develop and formulate new research questions. As and when the ethical issues related to AI and ML in health financing are addressed and coupled with feasible regulatory options, there can be accelerated progress towards UHC.

# References

Alami H, Rivard L, Lehoux P, Hoffman SJ, Cadeddu SBM, Savoldelli M et al. (2020). Artificial intelligence in health care: laying the foundation for responsible, sustainable, and inclusive innovation in low- and middle-income countries. Glob Health. 16(1):52.

Andini M, Ciani E, de Blasio G, D'Ignazio A, Salvestrini V (2018). Targeting with machine learning: an application to a tax rebate program in Italy. J EconB Behav Organ. 156:86–102.

Araújo FHD, Santana AM, de A Santos Neto P (2016). Using machine learning to support healthcare professionals in making preauthorisation decisions. Int J Med Inform. 94:1–7.

Ashrafian H, Darzi A (2018). Transforming health policy through machine learning. PLoS Medicine. 15(11):e1002692.

Chen H, Xing J, Yang X, Zhan K (2021). Heterogeneous effects of health insurance on rural children's health in China: a causal machine learning approach. Int J Environ Res Public Health. 18(18):9616.

Cinaroglu S (2020). Modelling unbalanced catastrophic health expenditure data by using machine-learning methods. Intell Syst Account Finance Manag. 27(4):168–81.

Davis S, Nourazari S, Granovsky R, Fard N (2021). Predicting a need for financial assistance in emergency department care. Healthcare (Basel). 9(5):556.

Donnenberg NS, Hernandez I, Normolle DP (2021). Determining the prevalence and risk factors for prescription drug unaffordability. Res Social Adm Pharm. Epub 2021 Jun 29. doi:10.1016/j.sapharm.2021.06.019.

dos Santos JRR, Dias CM, Filho AC (2021). Machine learning and national health data to improve evidence: finding segmentation in individuals without private insurance. Health Policy Technol. 10(1):79–86.

Dua P, Bais S (2014). Supervised learning methods for fraud detection in healthcare insurance. Machine learning in healthcare informatics. Heidelberg: Springer; 261–85.

Ekina T, Ieva F, Ruggeri F, Soyer R (2013). Application of Bayesian methods in detection of healthcare fraud. Chem Eng Trans. 33:81–9.

García-Centeno M-C, Mínguez-Salido R, del Pozo-Rubio R (2021). The classification of profiles of financial catastrophe caused by out-of-pocket payments: a methodological approach. Mathematics (Basel). 9(11):1170.

Gartner D, Kolisch R, Neill DB, Padman R (2015). Machine learning approaches for early DRG classification and resource allocation. INFORMS J Comput. 27(4):718–34.

Grover D, Bauhoff S, Friedman J (2019). Using supervised learning to select audit targets in performance-based financing in health: an example from Zambia. PloS One. 14(1):e0211262.

Hanson K, Brikci N, Erlangga D, Alebachew A, De Allegri M, Balabanova D et al. (2022). The Lancet Global Health Commission on financing primary health care: putting people at the centre. Lancet Glob Health. 10(5):e715–72.

Hassan AKI, Abraham A (2013). Computational intelligence models for insurance fraud detection: a review of a decade of research. Journal of Network and Innovative Computing. 1(2013):341–7.

Hileman G, Steele S (2016). Accuracy of claims-based risk scoring models. Schaumburg (IL): Society of Actuaries (https://www.soa.org/globalassets/assets/Files/Research/research-2016-accuracy-claims-based-risk-scoring-models.pdf, accessed 19 December 2022).

Ho CWL, Ali J, Caals K (2020). Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance. Bull World Health Organ. 98(4):263–9.

Islam MM, Li GH, Poly TN, Li YJ (2021). DeepDRG: performance of artificial intelligence model for real-time prediction of diagnosis-related groups. Healthcare (Basel). 9(12):1632.

Joedicke AM, Zellweger U, Tomka IT, Neuer T, Curkovic I, Roos M et al. (2019). Prediction of health care expenditure increase: how does pharmacotherapy contribute? BMC Health Serv Res. 19(1):953.

Joudaki H, Rashidian A, Minaei-Bidgoli B, Mahmoodi M, Gerail B, Nasiri M et al. (2016). Improving fraud and abuse detection in general physician claims: a data mining study. Int J Health Policy Manag. 5(3):165–72.

Kan HJ, Kharrazi H, Chang H-Y, Bodycombe D, Lemke K, Weiner JP (2019). Exploring the use of machine learning for risk adjustment: a comparison of standard and penalized linear regression models in predicting health care costs in older adults. PloS One. 14(3):e0213258.

Kasy M. (2018). Optimal taxation and insurance using machine learning — sufficient statistics and beyond. J Public Econ. 167:205–19.

Kirlidog M, Asuk C (2012). A fraud detection approach with data mining in health insurance. Procedia Soc Behav Sci. 62:989–94.

Kose I, Gokturk M, Kilic K (2015). An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance. Applied Soft Computing. 36:283–99.

Kreif N, Mirelman A, Moreno Serra R, Hidayat T, DiazOrdaz K, Suhrcke M (2020). Who benefits from health insurance? Uncovering heterogeneous policy impacts using causal machine learning [Working Paper]. York: University of York (https://eprints.whiterose.ac.uk/166393/1/CHERP173_health_insurance_causal_machine_learning.pdf, accessed July 2022).

Kriegeskorte N, Golan T (2019). Neural network models and deep learning. Current Biology. 29(7):R231–6.

Kumar M, Ghani R, Mei Z. Data mining to predict and prevent errors in health insurance claims processing. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington (DC), July 2010:65–74. New York (NY): Association for Computing Machinery (https://dl.acm.org/doi/10.1145/1835804.1835816, accessed 19 December 2022).

Liu J, Capurro D, Nguyen A, Verspoor K (2021a). Early prediction of diagnostic-related groups and estimation of hospital cost by processing clinical notes. NPJ Digit Med. 4(1):103.

Liu X, Fang C, Wu C, Yu J, Zhao Q (2021b). DRG grouping by machine learning: from expert-oriented to data-based method. BMC Med Inform Decision Making. 21(1):312.

Lu F, Boritz JE (2005). Detecting fraud in health insurance data: learning to model incomplete Benford's Law distributions. 16th European Conference on Machine Learning, Porto, Portugal, 3–7 October 2005 (https://link.springer.com/chapter/10.1007/11564096_63, accessed 19 December 2022). Proceedings. Heidelberg: Springer; 633–40.

Matloob I, Khan SA, Hussain F, Butt WH, Rukaiya R, Khalique F (2021). Need-based and optimized health insurance package using clustering algorithm. Appl Sci. 11(18):8478.

McGuire TG, Zink AL, Rose S (2021) Improving the performance of risk adjustment systems: constrained regressions, reinsurance, and variable selection. Am J Health Econ. 7(4).

Mrazek M, O'Neill F (2020). Artificial intelligence and healthcare in emerging markets. EM Compass: fresh ideas about business in emerging markets. Washington (DC): The World Bank (https://openknowledge.worldbank.org/handle/10986/34855, accessed 8 July 2022).

Mumtaz Z, Whiteford P (2021). Machine learning based approach for sustainable social protection policies in developing societies. Mob Netw Appl. 26(1:159–73.

Muremyi R, Haughton D, Kabano I, Niragire F (2020). Prediction of out-of-pocket health expenditures in Rwanda using machine learning techniques. Pan Afr Med J. 37:357.

Ng SHX, Rahman N, Ang IYH, Sridharan S, Ramachandran S, Wang DD et al. (2020). Characterising and predicting persistent high-cost utilisers in healthcare: a retrospective cohort study in Singapore. BMJ Open. 10(1):e031622.

NHA (2019) Beneficiary identification guidelines: Ayushman Bharat – Pradhan Mantri Jan Arogya Yojana (AB PM-JAY). New Delhi: India National Health Authority. (https://ayushmanup.in/admin/Clients/Doc/85_Guidelines-on-Process-of-Beneficiary-Identification.pdf, accessed 8 July 2022).

Nichol AA, Batten JN, Halley MC, Axelrod JK, Sankar PL, Cho MK (2021). A typology of existing machine learning-based predictive analytic tools focused on reducing costs and improving quality in health care: systematic search and content analysis. J Med Internet Res. 23(6):e26391.

Nomura Y, Ishii Y, Chiba Y, Suzuki S, Suzuki A, Suzuki S et al. (2021). Does last year's cost predict the present cost? An application of machine leaning for the Japanese Area-Basis Public Health Insurance Database. Int J Environ Res Public Health. 18(2):565.

OECD (2019). Recommendation of the Council on Artificial Intelligence (OECD Legal Instruments: OECD/LEGAL/O449). Paris: Organisation for Economic Co-operation and Development (https://www.fsmb.org/siteassets/artificial-intelligence/pdfs/oecd-recommendation-on-ai-en.pdf, accessed 19 December 2022).

OECD (2020). Trustworthy artificial intelligence in health. Background paper for the G20 AI Dialogue, Digital Economy Task Force (Saudi Arabia, 1–2 April 2020). Paris: Organisation for Economic Co-operation and Development (https://www.oecd.org/health/trustworthy-artificial-intelligence-in-health.pdf, accessed 19 December 2022).

OECD (2022). Towards an integrated health information system in Korea. Paris: Organisation for Economic Co-operation and Development (https://www.oecd-ilibrary.org/social-issues-migration-health/towards-an-integrated-health-information-system-in-korea_c4e6c88d-en, accessed 19 December 2022).

Ortega P, Figueroa C, Ruz G (2006). A medical claim fraud/abuse detection system based on data mining: a case study in Chile (https://www.adaptiveagents.org/lib/exe/fetch.php?media=papers/frauddetection.pdf, accessed 19 December 2022).

Panch T, Mattie H, Atun R (2019). Artificial intelligence and algorithmic bias: implications for health systems. Journal of global health, 9(2):010318.

Panesar A, Panesar H (2020). Artificial intelligence and machine learning in global healthcare. In: Haring R, Kickbusch I, Ganten D, Moeti M, editors. Handbook of global health. Cham: Springer International Publishing; 1–39 (https://link.springer.com/referencework/10.1007/978-3-030-45009-0, accessed 19 December 2022).

Rose S (2016). A machine learning framework for plan payment risk adjustment, Health services research. 51(6):2358–74.

Rose S (2018). Robust machine learning variable importance analyses of medical conditions for health care spending. Health Serv Res. 53(5):3836–54.

Rose S, Bergquist SL, Layton TJ. (2017). Computational health economics for identification of unprofitable health care enrollees. Biostatistics (Oxford). 18(4):682–94.

Shin H, Park H, Lee J, Jhee WC (2012). A scoring model to detect abusive billing patterns in health insurance claims. Expert Systems with Applications. 39(8):7441–50.

Sowah RA, Kuuboore M, Ofoli A, Kwofie S, Asiedu L, Koumadi KM et al. (2019). Decision support system (DSS) for fraud detection in health insurance claims using genetic support vector machines (GSVMs). Journal of Engineering (Cairo). 2019:1–19.

Sun H, Xiao J, Zhu W, He Y, Zhang S, Xu X et al. (2020). Medical knowledge graph to enhance fraud, waste, and abuse detection on claim data: model development and performance evaluation. JMIR Med Inform. 8(7):e17653.

Thesmar D, Sraer D, Pinheiro L, Dadson N, Veliche R, Greenberg P (2019). Combining the power of artificial intelligence with the richness of healthcare claims data: opportunities and challenges. PharmacoEconomics. 37(6):745–52.

UNESCO (2021). Report of the Social and Human Sciences Commission (SHS). Paris: United Nations Educational, Scientific and Cultural Organization (https://unesdoc.unesco.org/ark:/48223/pf0000379920, accessed 19 December 2022).

Usama M, Qadir J, Raza A, Arif H, Yau KA, Elkhatib Y et al. (2019). Unsupervised machine learning for networking: techniques, applications and research challenges. IEEE Access. 7:65579–615 (https://ieeexplore.ieee.org/document/8713992, accessed 19 December 2022).

Vimont A, Leleu H, Durand-Zaleski I (2022). Machine learning versus regression modelling in predicting individual healthcare costs from a representative sample of the nationwide claims database in France. Eur J Health Econ. 23(2):211–23 (the online pre-publication was found in 2021).

Waghade SS, Karandikar AM (2018). A comprehensive study of healthcare fraud detection based on machine learning. Int J Appl Eng Res. 13(6):4175–8.

WHO (2021a). Global strategy on digital health 2020–2025. Geneva: World Health Organization (https://apps.who.int/iris/handle/10665/344249, accessed 19 December 2022).

WHO (2021b). Digital technologies for health financing: what are the benefits and risks for UHC? Some initial reflections. (Health financing working paper, No. 19). Geneva: World Health Organization (https://apps.who.int/iris/handle/10665/343505, accessed 12 June 2022).

WHO (2021c). Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization (https://apps.who.int/iris/handle/10665/341996, accessed 19 December 2022).

WHO (2022). The use of digital technologies to support the identification of poor and vulnerable population groups for health coverage schemes. Insights from Cambodia, India and Rwanda. Geneva: World Health Organization (https://apps.who.int/iris/handle/10665/365368, accessed 19 December 2022).

WHO (forthcoming). Digital claims management in Estonia. A lever for strategic purchasing. (WHO technical report). Geneva: World Health Organization.

Xie Y, Neubauer S, Schreier G, Redmond SJ, Lovell NH (2015). Impact of hierarchies of clinical codes on predicting future days in hospital. Annu Int Conf IEEE Eng Med Biol Soc 2015. 2015:6852–5.

Yang C, Delcher C, Shenkman E, Ranka S (2018). Machine learning approaches for predicting high cost high need patient expenditures in health care. Biomedical engineering online. 17(Suppl 1):131–131.

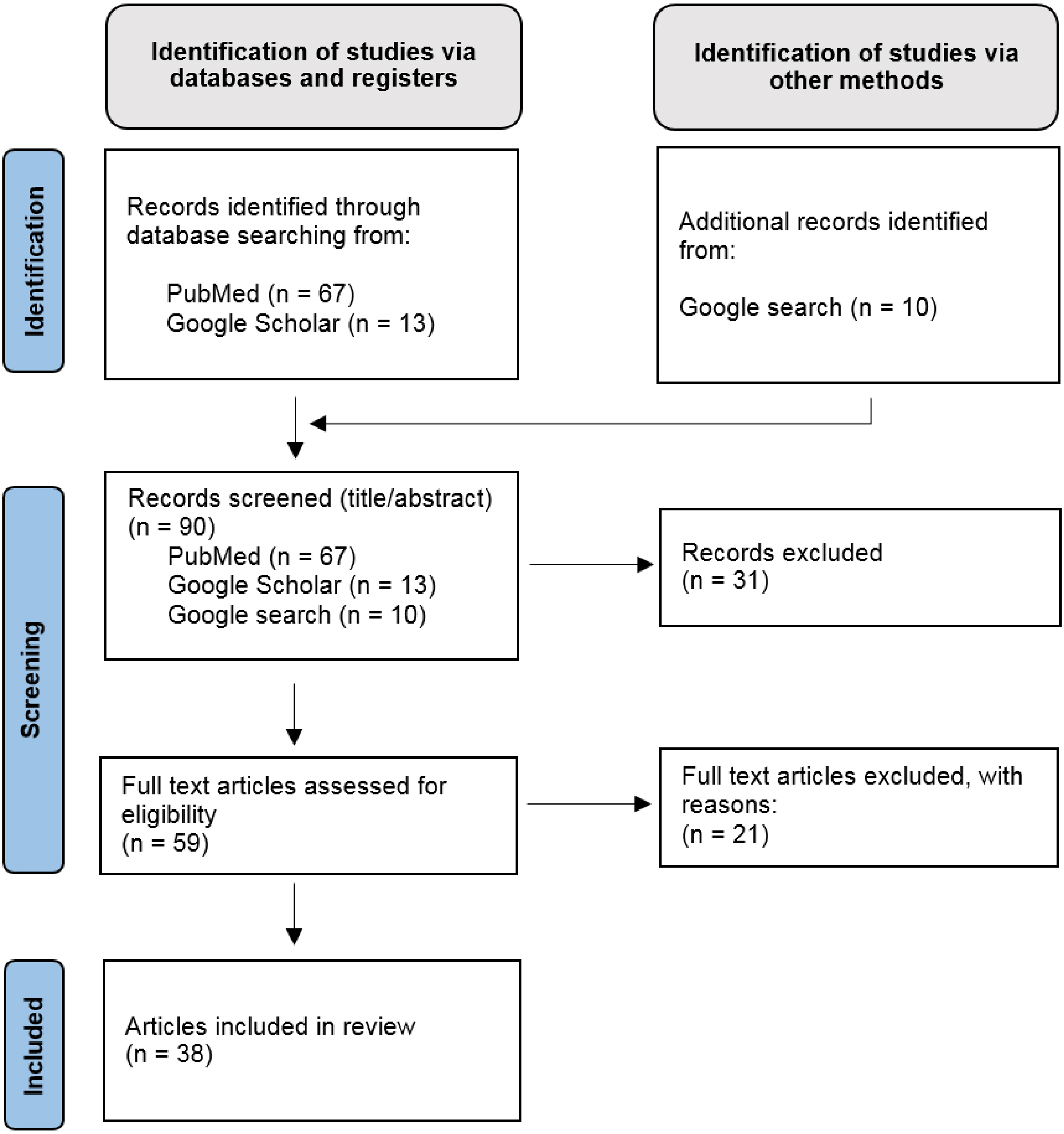
# Annexes

## Annex 1. PubMed search terms and strategy

| Search term (1)* | Search term (2)* |
|---|---|
| health financing OR<br>healthcare financing OR<br>revenue raising OR<br>pooling OR purchasing OR<br>health benefit package OR<br>health benefits OR<br>health insurance | artificial intelligence OR<br>machine learning OR<br>big data |
| (1) AND (2) | |

* These terms were equally used for the search in Google Scholar.

# Annex 2. Rapid literature review – Flowchart



# Annex 3. Summary table of papers included in this rapid literature review

| Author and year | Country *Type of data* | Type of paper | Type of AI/ML *Type of algorithm* | Purpose and findings |
|---|---|---|---|---|
| PREDICTION OF HEALTH EXPENDITURE | | | | |
| Joedicke et al. (2019) | Switzerland<br>*Compulsory private health insurance data* | Research (modelling) | Supervised ML<br>– classification<br>– variable selection<br>*(Logistic regression, boosted decision trees, neural networks)* | The objective of this paper is to predict whether health-care costs of individual patients will increase or decrease in the subsequent year, and to identify factors contributing to this prediction, with a particular focus on the role of pharmacotherapy. Using 2014–2015 Swiss health insurance claims data, the number of features is reduced from 747 to 36 with only a 0.5% loss in accuracy. Three algorithms are tested, of which the "boosted decision tree" model performs best. |
| Kan et al. (2019) | USA<br>*Compulsory private health insurance data (Medicare Advantage)* | Research (modelling) | Supervised ML<br>– regression<br>*(OLS regression, penalized linear regression, Lasso)* | This study conducts an in-depth comparison of various regression techniques to predict future health-care costs in older adults, using Medicare data from the USA. Penalized regression shows better performance than "ordinary least square" (OLS) regression in predicting health-care costs. Health-care insurers, providers and policy-makers may benefit from adopting penalized regression such as Lasso regression for cost prediction. |
| Ng et al. (2020) | Singapore<br>*Hospital data* | Research (modelling) | Supervised ML<br>– classification<br>– variable selection<br>*(Penalized regression, support vector machine, XGBoost)* | This paper aims to stratify a population of health-care utilizers into persistent high-cost utilizers, transient high-cost utilizers and non-high-cost utilizers, using inpatient and outpatient data from a hospital in Singapore. A model based on an extreme gradient boosting (XGBoost) algorithm outperforms two other algorithms and can predict at the end of one year whether a patient will continue to be a high utilizer in the next two years. |
| Nichol et al. (2021) | USA | Literature Review | Review of products and tools that use machine learning–based predictive analytics in health-care settings | This is a systematic literature review of studies and products that use ML–based predictive analytics in health-care settings. It identifies 38 "products" that predict utilization and/or costs. Common use cases include financial risk stratification of populations and the prediction of which individuals will become high(-cost) utilizers. Many algorithms have not been rigorously tested, and little is known about their comparison with other predictive analytics or clinical judgement. The paper also points to the underlying ethical tension when efficiency is improved by reducing cost with possible negative effects on quality. |
| Nomura et al. (2021) | Japan<br>*Public health insurance data* | Research (modelling) | Supervised ML<br>– regression/prediction<br>*(Neural network, support vector machine, gradient boosting machine)* | This paper analyses public health insurance data from Japan to predict the annual medical cost based on the medical cost and the dental health-care cost of the previous year, comparing three models. A neural network performs the best: medical cost proportionally depends on the medical cost of the previous year, while dental health-care cost of the previous year has a very small reducing effect on the overall medical cost. |

| Author and year | Country *Type of data* | Type of paper | Type of AI/ML *Type of algorithm* | Purpose and findings |
|---|---|---|---|---|
| PREDICTION OF HEALTH EXPENDITURE | | | | |
| Rose (2018) | USA<br>*Voluntary health insurance data* | Research (modelling) | Supervised ML<br>– regression/prediction<br>*(non-parametric targeted learning methods versus parametric regression)* | This study evaluates the average incremental annual cost related to 26 of the most prevalent medical conditions. The performance of a nonparametric targeted learning framework, which incorporates ensemble ML, is compared with traditional parametric regression. It concludes that the existing literature may be underestimating the spending contributions of specific medical conditions such as major depressive and bipolar disorders and chronic hepatitis. |
| Vimont, Leleu and Durand-Zaleski (2021) | France<br>*Public health insurance data* | Research (modelling) | Supervised ML<br>– regression/prediction<br>– variable selection<br>*(Random forest, neural network, generalized linear model)* | This paper compares the performance of a simple neural network and random forest to a generalized linear model (GLM) for the prediction of medical cost at the individual level, using a representative sample of all medical claims data for 2015 in France. It finds that random forest in most cases outperforms the other two models, while GLM is suited to understand the contribution of specific variables as predictors. This can help to develop better risk adjustment models (under competitive health insurance) or realize more efficient and equitable allocation of resources (under non-competitive insurance). |
| Xie et al. (2015) | Australia<br>*Voluntary health insurance data* | Research (modelling) | Supervised ML<br>– classification<br>– regression<br>*(Bagged regression trees)* | This methodological study tests whether the way in which data (i.e. clinical codes) are hierarchically structured can be utilized to improve their power to predict a patient's future days in hospital. The main finding from this study is that different hierarchies of clinical codes do not have a significant impact on predictive power. |
| Yang et al. (2018) | USA<br>*Public health insurance data* | Research (modelling) | Supervised ML<br>– regression/prediction<br>*(OLS regression, Lasso, gradient boosting machine, recurrent neural network)* | Looking at the Medicaid programme of Texas, this paper tests four predictive models to forecast expenditures, especially for "high-cost, high-need" patients: ordinary least squares linear regression, regularized regression (Lasso), gradient boosting machine, and recurrent neural networks (a deep learning approach). It is found that there is significant temporal correlation in health-care expenditures. The recurrent neural network demonstrates the highest prediction accuracy. |
| RISK SCORING: RISK ADJUSTMENT VERSUS RISK SELECTION | | | | |
| Hileman & Steele (2016) | USA<br>*Voluntary health insurance data* | Research (modelling) | Supervised ML<br>– regression/prediction<br>*(Random forest)* | This paper tested a random forest model, which had been trained with Medicare data, on its risk-scoring performance for voluntary health insurance data. The ML-based model uses diagnoses, pharmacy data and prior year cost to predict relative risk. Comparing its performance to the conventional risk-scoring models, it is concluded that the model may also have potential for commercial insurance populations. |
| Hileman & Steele (2016) | USA<br>*Voluntary health insurance data* | Research (modelling) | Supervised ML<br>– regression/prediction<br>*(Random forest)* | This paper tested a random forest model, which had been trained with Medicare data, on its risk-scoring performance for voluntary health insurance data. The ML-based model uses diagnoses, pharmacy data and prior year cost to predict relative risk. Comparing its performance to the conventional risk-scoring models, it is concluded that the model may also have potential for commercial insurance populations. |

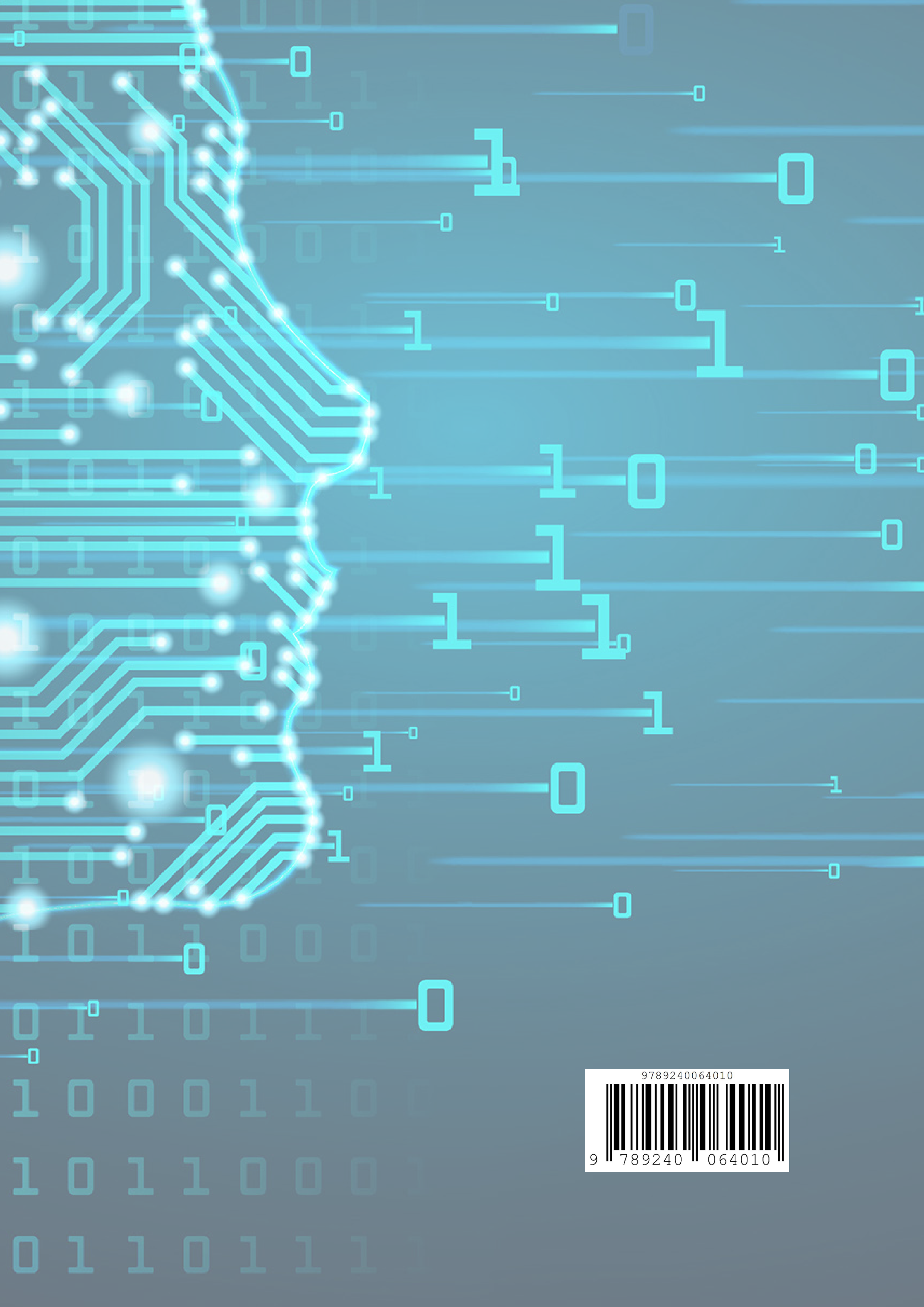
| **Author and year** | **Country** ***Type of data*** | **Type of paper** | **Type of AI/ML** ***Type of algorithm*** | **Purpose and findings** |
|---|---|---|---|---|
| RISK SCORING: RISK ADJUSTMENT VERSUS RISK SELECTION | | | | |
| McGuire, Zink & Rose (2021) | USA<br>*Compulsory private health insurance data* | Research (modelling) | Supervised ML<br>– regression<br>– variable selection<br>*(Random forest, Lasso penalized regression, etc.)* | The paper seeks to improve performance of risk adjustment models (health plan payment systems). It challenges the usual method of improving formulas by increasing the number of variables: with a reduced number of attributes, the authors claim performance is equal to or better than the current payment formula. Working with fewer variables reduces costs and limits opportunities for gaming, while a more uncertainty can be reinsured. |
| Mrazek & O'Neill (2020) | South Africa<br>*Voluntary health insurance data* | Topic overview (World Bank) | Various uses of AI/ML in general | The paper briefly describes the approach of Discovery, one of South Africa's largest private health insurance companies. Discovery applies ML to customer data accrued from supermarket purchases, activities in fitness firms and health-care utilization at facilities, in order to assess the level of customers' healthy behaviour. Customers are offered financial rewards, while the insurer reduces the average health risks of its member pool. |
| Rose (2016) | USA<br>*Voluntary health insurance data* | Research (modelling) | Supervised ML<br>– regression<br>– variable selection<br>*(Seven individual algorithms and one ensemble)* | The paper seeks to improve the risk adjustment mechanism among insurance schemes, testing several algorithms. An ensemble based on a weighted average of multiple algorithms is found to maintain much of the efficiency of a traditional larger formula and outperforms the individual algorithms, as well as classical regression. A simplified formula reduces incentives for increased coding intensity and upcoding. |
| Rose, Bergquist & Layton (2017) | USA<br>*Compulsory private health insurance data* | Research (modelling) | Supervised ML<br>– regression<br>– variable selection<br>*(Five individual algorithms and one ensemble)* | The paper assesses the prescription drug formulary as one specific component of health plan design that competing insurers in the United States Health Insurance Marketplace may use for risk selection. It demonstrates that an ensemble (a "super learner") is effective in extracting the relevant signal for this prediction problem, and that a small number of drug variables can be used to identify "unprofitable" enrollees. The paper calls for regulation to mitigate the risks. |
| CLAIMS MANAGEMENT AND FRAUD DETECTION | | | | |
| Araújo, Santana & de Santos Neto (2016) | Brazil<br>*Voluntary health insurance data* | Research (modelling) | Supervised ML<br>– classification<br>– variable selection<br>*(Random tree, naïve Bayes, support vector machine, nearest neighbor, one ensemble)* | This paper aims to test models that can support the process of pre-authorization of dental claims by professional reviewers in Brazil, so that less complex cases may be evaluated fully automatically. The number of attributes is reduced from 164 to 15, and 12 new attributes are created from existing data. The performance of classifiers (random tree, naïve Bayes, support vector machine and nearest neighbor is evaluated. The three classifiers that yielded superior results (random tree, support vector machine and nearest neighbor) are combined using an ensemble. The resulting decision support mechanism achieves a hit rate of above 96%. |

| Author and year | Country *Type of data* | Type of paper | Type of AI/ML *Type of algorithm* | Purpose and findings |
|---|---|---|---|---|
| CLAIMS MANAGEMENT AND FRAUD DETECTION | | | | |
| Ekin et al. (2013) | - *Simulated data* | Research (modelling) | Unsupervised ML – clustering *(Bayesian co-clustering)* | In this paper, application of Bayesian ideas in health-care fraud detection is presented, with a focus on identification of potentially fraudulent providers and beneficiaries who have unusual group memberships. Using simulated data, it is demonstrated that Bayesian co-clustering can help to identify potentially fraudulent providers and beneficiaries. |
| Grover, Bauhoff & Friedman (2019) | Zambia *Health centre performance and facility survey data* | Research (modelling) | Supervised ML – classification *(Naïve Bayes, logistic regression, support vector machines, random forest)* | This paper tests various ML techniques to identify health centres in Zambia (which receive performance-based payments) that would need to be targeted for verification in order to reduce costs. Four methods of random sampling and predictive modelling are compared, based on naïve Bayes, logistic regression, support vector machines and random forest. The latter outperforms the other ML techniques; its application could lead to substantial cost savings, by reducing both fraudulent claims and unnecessary verifications. |
| Joudaki et al. (2016) | Iran (Islamic Republic of) *Public health insurance data* | Research (modelling) | Unsupervised ML – clustering Supervised ML – discriminant analysis | The paper uses cluster analysis to scrutinize general physicians' prescription claims in the Islamic Republic of Iran. The model extracts a subset of claims or claimants for further assessment and scrutiny. 54% of general physicians are "suspects" of abusive behaviour and 2% are suspected of fraud. Discriminant analysis is used to assess the validity of the clustering approach, suggesting that the indicators demonstrate adequate performance in detection of physicians who were suspected of fraud (98%) and abuse (85%) in a new sample of data. |
| Kirlidog & Asuk (2012) | Türkiye *Voluntary health insurance data* | Research (modelling) | Supervised ML – classification *(Support vector machine)* | This paper discusses different types of fraud in health insurance in Türkiye, using data from a private health insurer to test a support vector machine algorithm to detect anomalies. Based on previous cases that are known or suspected to be fraudulent, the model calculates the likelihood or probability that each record is fraudulent. |
| Kose, Gokturk & Kilic (2015) | Türkiye *Voluntary health insurance data* | Research (modelling) | Semi-supervised ML – classification | The purpose of this study is to evaluate a novel framework to detect fraudulent and abusive cases independently from the actors and commodities involved in the claims. It applies interactive ML, that allows for the incorporation of expert knowledge in a setting of unsupervised learning. The proposed framework is evaluated with real-life data for six different abnormal behaviour types for prescriptions. The system significantly reduces the time requirements for the fact-finding process and is utilized to produce monthly reports that include abnormal behaviours for evaluation by the insurance company. |
| Kumar, Ghani & Mei (2010) | USA *Voluntary health insurance data* | Research (modelling) | Supervised ML – classification – variable selection *(Support vector machines)* | This paper presents a system to help classify health insurance claims that are potentially problematic or that require specific attention. Using data from a large health insurer, it claims a higher precision and sensitivity than existing approaches. Without clear substantiation, potential savings of US$ 15–25 million per year are claimed for a typical US health insurer because payment errors and work associated with reprocessing claims can be avoided. |

| Author and year | Country *Type of data* | Type of paper | Type of AI/ML *Type of algorithm* | Purpose and findings |
|---|---|---|---|---|
| CLAIMS MANAGEMENT AND FRAUD DETECTION | | | | |
| Liu X et al. (2021b) | China<br>*Hospital data (real and simulated)* | Research (modelling) | Supervised ML<br>– *classification*<br>*(five classification techniques)* | This study proposes a method of data-based grouping, based on ML algorithms and trained by real or simulated cases. It inherits the decision-making rules from the expert-oriented grouping and improves performance by incorporating continuous updates. Five classification algorithms are tested (naïve Bayes, support vector machine, classification and regression trees (CART), random forest, XGBoost). Some algorithms demonstrate equal or better performance than expert-based grouping, with the advantage of less subjectivity and low costs for designing / updating the system. |
| Lu & Boritz (2005) | Canada<br>*Voluntary health insurance data* | Research (modelling) | Unsupervised ML<br>– anomaly detection<br>*(Adaptive Benford algorithm)* | Using private insurance data, the paper describes a fraud detection model that utilizes an unsupervised learning approach to handle incomplete data. It is based on an enhanced version of "Benford's law" (a probability distribution of digits that can be used to detect fraud, for instance in accounting or in election data). It demonstrates improved precision over the traditional Benford approach in detecting anomalous data indicative of fraud. |
| Ortega, Figueroa & Ru (2006) | Chile<br>*Voluntary health insurance data* | Research (modelling) | Supervised ML<br>– classification<br>*(Multilayer perceptron neural networks)* | This paper uses private health insurance data from Chile, and proposes a fraud detection system, which uses one committee of multilayer perceptron neural networks for each one of the entities involved: medical claims, affiliates, medical professionals and employers. It finds a much higher detection rate than the insurance company had previously recorded, whereby cases are also being detected more than 6 months earlier. |
| Shin et al. (2012) | Republic of Korea<br>*Public health insurance data* | Research (modelling) | Supervised ML<br>– classification<br>*(Decision trees)* | This paper proposes a scoring model that detects outpatient clinics with abusive utilization patterns on the basis of profiling information extracted from electronic insurance claims. The model consists of: 1) scoring to quantify the degree of abusiveness; and 2) segmentation to categorize problematic providers with similar utilization patterns. The validation indicates that the proposed model is largely consistent with the manual detection techniques currently used to identify potential abusers. The proposed model allows for automation of abuse detection and is an opportunity to reduce costs. |
| Sowah et al. (2019) | Ghana<br>*Public health insurance data* | Research (modelling) | Supervised ML<br>– classification<br>*(Genetic algorithms, support vector machines)* | In Ghana, Sowah et al. (2019) combine so-called genetic algorithms and support vector machines (GSVMs) to solve the health insurance claim classification problem and eliminate fraudulent claims. It evaluates three GSVM classifiers and demonstrates improved accuracy of classification in legitimate and fraudulent claims, while also significantly reducing the computational time needed for claims processing. |
| Sun et al. (2020) | China<br>*Voluntary health insurance data* | Research (modelling) | Supervised ML<br>– classification<br>– natural language processing (NLP)<br>*(Knowledge graph, NLP, convolutional neural networks)* | The aim of this study is to develop an automated method to identify clinically suspect claims and to detect fraud, waste and abuse. Three knowledge graph reasoning rules are defined to identify three kinds of inappropriate diagnosis medications. To guarantee the quality of the medical knowledge graph, human experts are involved to review the entity and relationships with lower confidence. The model is claimed to detect automatically 70% of the claims that are clinically suspect. The medical knowledge graph–based method provides good interpretability of the results. |

| Author and year | Country *Type of data* | Type of paper | Type of AI/ML *Type of algorithm* | Purpose and findings |
|---|---|---|---|---|
| CLAIMS MANAGEMENT AND FRAUD DETECTION | | | | |
| Cinaroglu (2020) | Türkiye<br>*Household survey data* | Research (modelling) | Supervised ML<br>– classification<br>*(Random forest, logistic regression)* | This study aims to compare the performance of two classification algorithms (logistic regression and random forest) for the prediction of households that will face catastrophic health expenditure. Data were derived from a 2012 nationally representative household budget survey. Random forest exhibits superior performance in identifying such households. Household head characteristics (such as an age of 65 years or over) prove to be the best predictors of catastrophic health expenditure. |
| Davis et al. (2021) | USA<br>*Billing data (public and private health insurance members)* | Research (modelling) | Supervised ML<br>– classification<br>– variable selection<br>*(Decision tree, logistic regression, random forest)* | The goal of this study is to predict the likelihood of patients paying the balance of their emergency department visit bill within 90 days of receipt, in order to identify patients in need of government financial assistance. Patient billing records from the United States are used and three algorithms are compared: decision tree, logistic regression and random forest. The decision tree is found to be most accurate, correctly predicting 87% of the unsuccessful payments. |
| Donnenberg, Hernandez & Normolle (2021) | USA<br>*Household survey data* | Research (modelling) | Supervised ML<br>– regression<br>– variable selection<br>*(Lasso selection, logistic regression)* | This study aims to identify factors associated with the inability of patients to afford prescription medication, using data from the 2017 National Health Interview Survey. Covariates include demographic data, medical history and social attitudes. Lasso selection and logistic regression are used to identify predictors of cost-based prescription refusal. Eight main factors are identified, most importantly low income and a history of chronic medical conditions. |
| García-Centeno, Mínguez-Salido & del Pozo-Rubio (2021) | Spain<br>*Household survey data* | Research (modelling) | Supervised ML<br>– classification<br>*(Seven different classification techniques)* | The aim of this work is to predict which households are at risk of catastrophic expenditure as a result of out-of-pocket expenditure in long-term care in Spain, based on data from the Spanish Disability and Dependency Survey (SDDS) of 2008. A series of different classification techniques is tested. All models perform well, with more complex models – support vector machines and random forest – outperforming the simpler ones. |
| IDENTIFICATION OF HOUSEHOLDS FOR TARGETED POLICIES | | | | |
| Mumtaz & Whiteford (2021) | Pakistan<br>*Household survey data* | Research (modelling) | Unsupervised ML<br>– clustering<br>*(K-means clustering)* | This paper uses socioeconomic survey data from Pakistan and K-means clustering to cluster households according to their need for government assistance and to identify geographical differences. Four clusters of households are found, exhibiting common patterns related to their risk exposure in terms of loss of income and property, unemployment, disaster and disease. Specific rural areas are identified where households are in urgent need of social protection interventions. This kind of unsupervised learning could well be used in countries with data constraints. |
| Muremyi et al. (2020) | Rwanda<br>*Household survey data* | Research (modelling) | Supervised ML<br>– regression/prediction<br>– variable selection<br>*(Five different algorithms)* | By using data from the 2018 Rwanda Integrated Living Conditions Surveys, this paper aims to identify the best ML technique to predict out-of-pocket health expenditure, comparing the performance of five different models. A Treenet model demonstrates the highest prediction accuracy. The total household consumption is found to be the most important predictor variable, and this is consistently true for all the algorithms tested. |

| Author and year | Country<br>*Type of data* | Type of paper | Type of AI/ML<br>*Type of algorithm* | Purpose and findings |
|---|---|---|---|---|
| HEALTH NEEDS-INFORMED BENEFIT PACKAGE DESIGN | | | | |
| Dos Santos, Dias & Filho (2021) | Portugal<br>*National health survey data* | Research (modelling) | Unsupervised ML<br>– clustering<br>*(Multiple correspondence analysis [MCA])* | This paper applies multiple correspondence analysis to data from a nationwide health survey in Portugal in order to identify and describe segments within the uninsured population. Three distinct clusters of uninsured people are found: 1) young and middle-aged, healthy individuals with a relatively high income and education (mostly male); 2) older individuals who perceive their own health as fair but have long-standing or chronic conditions (mostly female); and 3) older individuals with bad self-reported health and symptoms of depression. The specific characteristics of the needs of these groups could be used to inform benefits design and public policies to improve access to health care. |
| Kasy (2018) | USA<br>*Health insurance data (experimental)* | Research (modelling) | Supervised ML<br>– regression/prediction<br>*(Nonparametric Bayesian methods, Gaussian process priors)* | This study reuses data that were collected in the United States between 1974 and 1981 for a health insurance experiment (RAND experiment) and applies Gaussian process regression to optimize policy decisions, such as tax rates or co-insurance rates in health insurance. It studies the effect of the co-insurance rate on the level of health-care expenditure and provides explicit formulas for posterior expected social welfare. Compared to previous studies which had used the same data, but had applied traditional statistical approaches, a much lower optimum co-insurance rate is found, namely 18% versus 50%. |
| Matloob et al. (2021) | Pakistan<br>*Voluntary health insurance data* | Research (modelling) | Unsupervised ML<br>– clustering<br>*(K-means clustering)* | The purpose of this paper is to inform the design of needs-based health insurance benefit packages based on the health-care records of individual employees, as opposed to the current practice of determining benefits on the basis of an employee's hierarchical position. Using five years of insurance data of hospital employees with group health insurance in Pakistan, this paper first applies a K-means clustering methodology to historical medical records. Subsequently, medical benefit optimization is achieved by applying a probability distribution model to the employees' insurance records. On average, a different categorization would result in a 25% increase in benefit amounts per employee. |
| HEALTH FINANCING-RELATED DETERMINANTS OF HEALTH SERVICE UTILIZATION | | | | |
| Chen et al. (2021) | China<br>*Household survey data, inter alia on public health insurance* | Research (policy analysis) | Supervised ML<br>– regression/prediction<br>– variable selection<br>*(Causal forest)* | This paper investigates the impact of Urban and Rural Resident Basic Medical Insurance (URRBMI) on the health of preschool and school-age children in rural China. Propensity score matching is applied, together with ML (causal forests) to evaluate impact. URRBMI significantly improves the health status of preschool children, while there is no significant effect for school-age children. For preschool children, disadvantaged mothers benefit more. The study demonstrates the power of causal forest to uncover heterogeneity in effects, providing policy-makers with valuable information for policy design. |
| Kreif et al. (2020) | Indonesia<br>*Household survey data, inter alia on public health insurance* | Research (policy analysis) | Supervised ML<br>– regression/prediction<br>– variable selection<br>*(Causal forest)* | This paper evaluates the impact of two health insurance schemes in Indonesia (subsidized and contributory). Using causal forests, beneficial impact is found for contributory health insurance (i.e. a significantly higher skilled birth attendance rate and reduced infant mortality) as compared to non-contributory or subsidized insurance. Significant heterogeneity is found in the impact of contributory health insurance: lower wealth quintiles benefit most. No significant heterogeneity is found for the subsidized scheme, even though it targeted vulnerable populations. The study demonstrates the power to uncover heterogeneity in impact, providing policy-makers with valuable information for programme design. |



9789240064010