Machine Learning in Health: A Systematic Review

Alessandra R. C. Padovam[1], Danilo Rodrigues Pereira[1], Romis Attux[1]

[1]Faculdade de Engenharia Elétrica e de Computação - Universidade Estadual de Campinas (UNICAMP)

Av. Albert Einstein, 400, Campinas - SP – Brazil CEP 13083-852

lecardos@unicamp.br, danilorp@unicamp.br, attux@unicamp.br

Abstract. *The aim of this study is to construct a taxonomy of Machine Learning (ML) techniques most commonly employed in health applications. The systematic review was performed between October 2023 and November 2024, and comprised articles published from 2013 to 2023. The search for relevant references was done on Pubmed and MEDLINE databases, and the screening articles were analysed by two reviewers independently through the Covidence platform. We identified 2,714 articles resulting in 65 health problems related to some ML technique and 8 types of learning. This research brings trends in use of ML for application in the field of medicine in general.*

1. Introduction

The last decade has witnessed remarkable developments in Artificial Intelligence (AI) methods, which have become essential to virtually all fields of modern engineering [Russell 2020]. These methods are also becoming part of the everyday life of billions of people around the globe, helping them in countless tasks and even in creative / artistic endeavors. The field of Machine Learning (ML) can be understood as a branch of modern AI that, in simple terms, encompasses a variety of techniques that allow computational models to learn from data i.e. to automatically modify their parameters in response to the information content of a representative set of stimuli [Bishop 2006].

The impact of ML in the broad subject of Health has been remarkable [Saxena 2021][Siddiqui 2015]. Focusing on Healthcare , this study presents a systematic review of the literature on ML techniques that can be applied to solve commonly observed problems like detection of diseases, early diagnosis and medical image segmentation. The design of modern ML methods takes place within an environment in which large amounts of data can be produced by an Internet of Medical Things (IoMT) interconnecting wearable devices - like smartwatches and glucose monitors -, health monitoring systems, implantable devices, smart pills etc. Moreover, these methods benefit from state-of-the-art solutions in radiographics, Magnetic Resonance Imaging (MRI) and Computerized Tomography (CT) [Topol 2019].

To better understand the structure of the health connectedness ecosystem, it is important to build a clear view on fundamental ML concepts, like learning paradigms, evaluation metrics, models etc., as well as on how to apply these models to typical health-related tasks. We believe this review will be of assistance to readers that wish to understand how these concepts have been presently employed in scenarios that are shaping the future of health in this new AI-based digital world. The paper is structured as follows: In section 2, we present the methodology for selection and analysis of studies describing the criteria and resources used. In section 3 we approach the main findings in the review. Finally in section 4 we discuss the number of cancer-related studies that use ML techniques. Additional information about results found were described in appendix I and II.

2. Methodology

This work is a systematic review of literature based on Preferred Items for Systematic Reviews and Meta Analyses (PRISMA) checklists [Moher 2015] and Cochrane guideline [Higgins 2019]. This systematic review is descriptive in nature and was conducted with the aim of characterizing the current state of the literature on ML in health, without applying methods to assess risk of bias or quality of evidence. The review protocol was prospectively registered with International Platform of Registered Systematic Review and Meta-analysis Protocols (INPLASY) - INPLASY202540097 https://inplasy.com/inplasy-2025-4-0097/

This research aims to answer this fundamental question: What are the machine learning methods most commonly used for solving health-related problems? This research question (RQ) leads to some specific objectives like: (a) to identify and to classify ML techniques employed in health problems; (b) to identify diseases and health conditions and to map health application areas; (c) to identify the most relevant evaluation metrics and factors for analyzing ML methods in health applications; (d) to highlight future directions of researches on healthcare applying ML.

The PICO related this aim as follows: **Population**: Within the scope of our study, the population included research involving adults and children, of any age and gender, in a variety of health contexts. **Intervention**: Articles that have used machine learning technique and belong to the health application research for any pathology or disease; **Comparisons**: In the context of solutions to health problems, the comparison between: different types of machine learning used; different data sets; different metrics for evaluating ML models; the performance of ML algorithms versus conventional or manual methods; **Outcomes**: ML metrics or statistics for model evaluation; Impact of the ML solution in the medical field; Robustness and generalizability of the model; Availability of a final product for use by healthcare professionals.

The literature search was performed through MEDLINE and PubMed databases focused on human studies. We performed a search - during the months of September and October 2023 - of articles published in the ten previous years (between January 1st, 2013 and October 25th, 2023). In the searches, the following English Medical Subject Headings (MeSH) terms were considered: “Machine Learning” and “Health Application”.

Review was screened by pairs (ARCP and DRP) independently by a second reviewer to minimize errors and every article. Using the Covidence platform, a systematic review was conducted. Covidence is a web-based collaboration software platform that streamlines the production of systematic and other literature reviews. [covidence.org 2024].

Following search strings used for the literature study:

On the Embase database we found 8314 papers and we, applied the following query: *('machine learning'/exp OR 'machine learning' OR (('machine'/exp OR machine)*

AND ('learning'/exp OR learning))) AND ('health application' OR (('health'/exp OR health) AND ('application'/exp OR application))) AND [english]/lim AND [humans]/lim AND [01-01-2003]/sd NOT [26-10-2023]/sd.

On the PubMed database we found 4815 papers and we, applied the following query: *(("machine learning"[MeSH Terms] OR ("machine"[All Fields] AND "learning"[All Fields]) OR "machine learning"[All Fields]) AND ("health"[MeSH Terms] OR "health"[All Fields] OR "health s"[All Fields] OR "healthful"[All Fields] OR "healthfulness"[All Fields] OR "healths"[All Fields]) AND ("applicabilities"[All Fields] OR "applicability"[All Fields] OR "application"[All Fields] OR "applications"[All Fields] OR "applicative"[All Fields])) AND ((humans[Filter]) AND (2013/1/1:2023/10/25[pdat]) AND (english[Filter]))*

Firstly, we created a project on the platform and uploaded the total number of articles extracted from the Embase and PubMed databases, totaling 13129 articles to be evaluated, of which 2386 were identified as duplicates and removed automatically by the tool (29 articles identified manually and 2357 identified by covidende). The covidence tool identified 10743 articles for screening phase during which the titles and abstracts were analysed by two reviewers resulting in 4197 irrelevant articles. The full text review was conducted in pairs too from 6546 studies assessed for eligibility, resulting in 3832 excluded articles. A total of 2714 articles that met the inclusion and exclusion criteria were represented in the PRISMA flowchart (Figure 1).

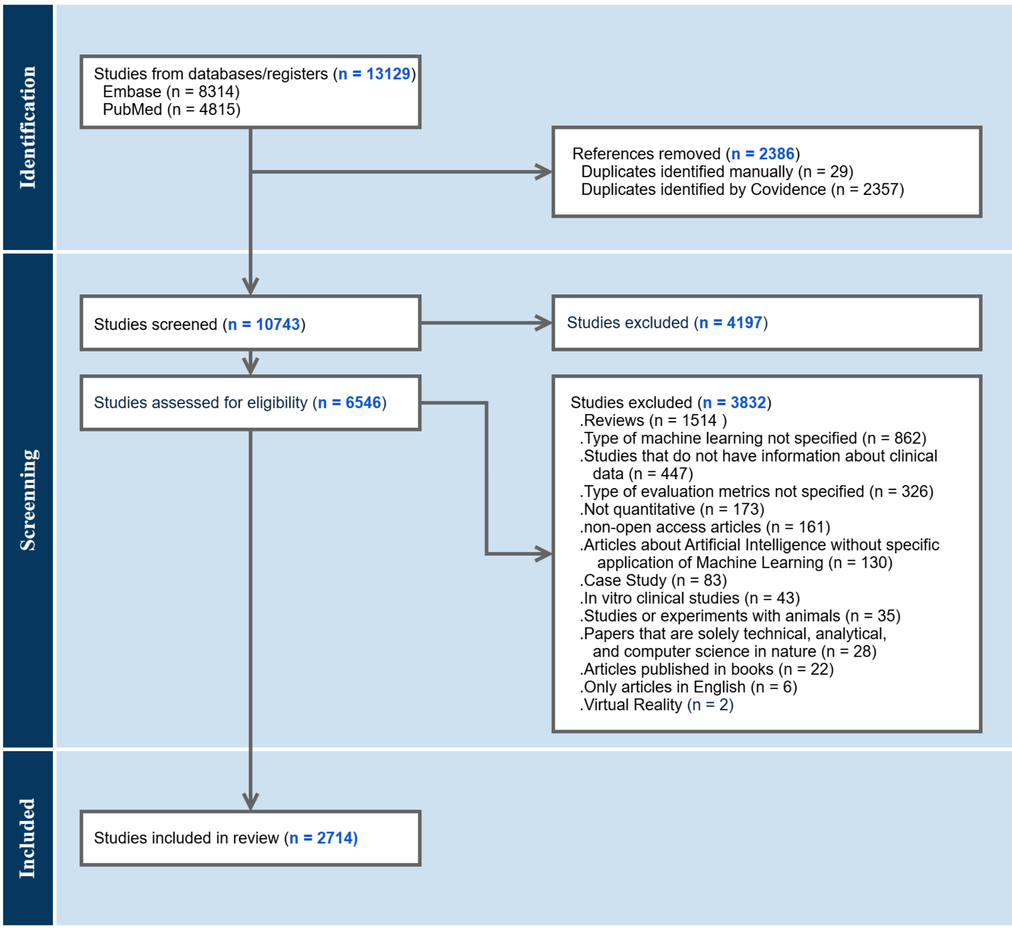

Figure 1: PRISMA flowchart of this systematic literature review on ML in health.

Disagreements between reviews during article screening were re-evaluated by one of the reviewers.

The following inclusion criteria (IC) were applied: a) Original Articles; b) Articles in English; c) Articles published in journals, conferences and workshops; d) Studies published between January 1st, 2013 and October 25th, 2023 in English; e) Studies focused on Machine Learning techniques for health applications.

The following exclusion criteria were applied: a) Studies or experiments with animals; b) Studies that do not have information about clinical data; c) Articles published in books; d) Articles about Artificial Intelligence without specific application of Machine Learning; e) Case Study; f) In vitro clinical studies; g) non-open access articles; h) Not quantitative; i) Only articles in English; j) Papers that are solely technical, analytical, and computer science in nature; k) Reviews; l) Type of evaluation metrics not specified; m) Type of machine learning not specified; n) Virtual Reality. It is important to highlight that we established review articles as an exclusion criterion to avoid redundancy of data and insurance that only original research were screened. We also excluded articles that mentioned IA only in broad terms, without specific application or evaluation ML methods.

After applying the inclusion and exclusion criteria in the screening of articles, 2714 studies were selected for data extraction. Due to the complexity of processing a large number of articles, it was necessary to perform a broad analysis to capture the general trends of the health areas, groups of pathologies, types of learning and metrics used to validate the models. Then from 2714 papers were submitted for analysis criterias like: aim of proposed model ML on study, pathology, pathology details, main health problem, learning type, type of ML technique employed, type of ML evaluation metric, population involved. The classification of pathologies mentioned in the screened articles was based on chapters of the International Classification of Diseases and Related Health Problems (ICD) 11th revision - Mortality and Morbidity Statistics (ICD-11-MMS) standard established by the World Health Organization [WHO] which is a consolidated global standard for health diagnostic information [Harrison 2021] . The wide analysis helped to identify the cut-off for a more detailed analysis work.

3. Results

This systematic review started on October 23rd, 2023 and finished on November 4th, 2024. In this section, we describe the results found for the research questions mentioned in the methodology. The table 1 shows the number of occurrences of eight ML paradigms among the selected articles. The table shows a prevalence of supervised over unsupervised methods and also indicates the relevance of neural network approaches, both in deep learning and classical formulations. It is important to point out that there is a certain conceptual intersection (for instance, between deep learning and supervised learning): our choice was based on an assessment of the aspect that seemed more pertinent in each work.

Table 1. Machine learning types extracted from included articles.

Machine Learning Type	Total
Supervised Learning	1419
Deep Learning	1008
Artificial Neural Networks	131
Unsupervised Learning	95
Supervised and Unsupervised Learning	37
Reinforcement Learning	12
Ensemble Model	6
Semi Supervised Learning	6

Figure 2 presents a word cloud representation that illustrates the frequency of occurrence of the pathologies found in the included articles - highlights words in larger sizes the pathologies found most frequently in articles and gradually reduces the size for less common terms. In total 65 pathologies were listed. Cancer being the disease found in the majority of studies - 488 articles. This result gives support to the consolidated indicators published by Global Cancer Observatory of the International Agency for Research in Cancer - World Health Organization, which provides comprehensive epidemiological data, including incidence, mortality and prevalence of cancer types [Ferlay 2021], [Ferlay 2024]. In the sequence of mostly disease lists were Cardiovascular Disease (CD) - 241 articles; Mental Disorder (MD) - 207 articles; Covid-19 - 139 articles; Endocrine Disease (ED) - 105 articles; Disorder of Bone (DB) - 100 articles; Neurodegenerative Disorder (NDD) - 95 articles, Pulmonary Disease (PD) - 95; Infectious Disease (ID) - 86 articles; Neurological disorder (ND) - 83 articles.

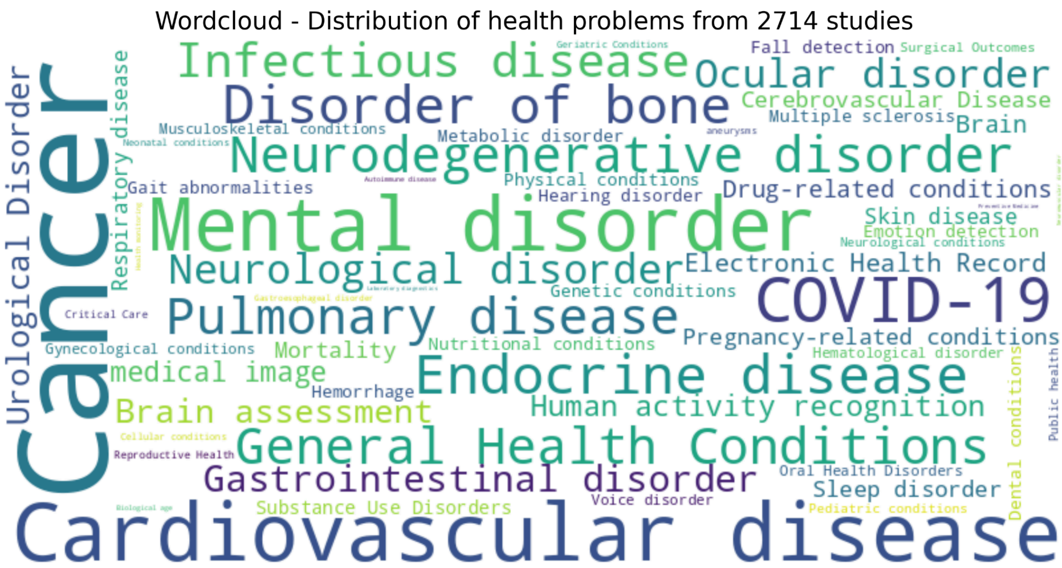

Figure 2 - Word Cloud Representation of Systematic Review about Health Problems Found from 2,714 Articles.

To provide a clearer view on the technical content of the 2714 selected works, we built two heat maps, Figures 3 and 4, found in the Appendices 1 and 2. In these maps, darker colors indicate a higher number of occurrences.

In the heat map shown in Figure 3, the papers are categorized according to ICD-11-MMS [Harrison 2021] by chapter number and title, resulting in a total of 2246 articles, with 41 diseases distributed over the x-axis and eight ML paradigms distributed over the y-axis, highlighting concentrations of warmer colors for 233 articles on Chapter 02 'Neoplasms' (Cancer) correlated with the SL, and 207 articles for Deep Learning (DL). Following this, there is a lower concentration of 153 articles on Chapter 8 'Disease of the Nervous System' (like Parkinson, Alzheimer) with SL and 73 for DL. Chapter 11 'Disease of the Circulatory System' (Cardiovascular Disease) with 126 for SL and 107 for DL, Chapter 06 'Mental, behavioural or neurodevelopmental disorders' (Mental Disorder) with 143 for SL, Chapter 05 'Endocrine, nutritional or metabolic diseases' with 89 for SL and finally Chapter 25 'International provisional assignment of new diseases of uncertain aetiology and emergency use' (COVID-19) with 65 articles for DL and 63 for SL.

The first heat map reveals the predominant use of DL in cancer studies due to its ability to handle complex data such as medical images, since the detection of cancer tumors is usually carried out through imaging exams. [Islam 2024]. Analyses of the nervous system are commonly based on electroencephalograms and magnetic resonance imaging explored by Recurrent Neural Networks. The circulatory system with a high index revealed in the heat map indicates the use of labeled data such as laboratory tests or electrocardiograms.

In the second heat map, we categorized articles that did not mention a specific pathology that fit the ICD classification. These articles were termed as General Health Problems (GHP), resulting in 468 articles where the areas of low article density are represented, with 25 GHP on the x-axis and 8 types of ML on the y-axis. Thus, the highest frequencies found were respectively for SL: General Health Conditions, Human Activity Recognition, Mortality, Brain Assessment, Drug-related Conditions; for DL: Medical Image, General Health Conditions, Brain Assessment. The second heat map also indicates the extensive application of DL for medical imaging investigation and also strong presence of SL technique.

4. Discussion

Inspired by [Triantafyllidis, 2019], which proposed to examine applications of machine learning in real-life digital health interventions, this article performs a systematic review with the aim of identifying the most commonly applied ML methods for healthcare problems. Our approach aims to map the application of methods in different clinical contexts, highlighting trends and gaps in the literature. This initial systematic review provided an overview of the research articles landscape that address or implement ML applied to solve different health problems and which will guide us to the second stage to filter the extraction targeted for the more detailed analysis in the combination of Supervised Learning (SL) and Unsupervised Learning (UL).

The synthesis of our findings revealed the highest concentration of research focused on Cancer research correlated with ML, totaling a rate of 17% percent of articles, double the rate found for cardiovascular diseases and mental illnesses, with 9% and 8% respectively. From this indicator of cancer research, it is notable that researchers are highly interested in this important disease. This study suggests that the highlighted types of ML have good adherence to the implementation of algorithms as a solution to health problems for Cancer, cardiovascular disease, corroborating with the studies [Kourou 2021], [Quer 2021].

About the types of learning we conclude that SL, due to the availability of labeled data, facilitates the diagnosis and prognosis of the pathology. UL clustering is useful for discovering heterogeneous patterns in unlabeled data, in addition to the preparation and reduction of noisy data made possible by Principal Component Analysis (PCA) , by projecting a high-dimensional data set into a smaller feature space, PCA also minimizes or completely eliminates common problems such as multicollinearity and overfitting. DL, on the other hand, favors the processing and analysis of medical images, due to the ability to extract features from complex data.

In clinical practice, the application of ML as an aid to diagnostic accuracy can be significant for improving patient response to treatment as well as personalizing treatments. The findings of this review highlight the potential use of ML in clinical decision support systems and the formulating public policies focused on digital health. In conclusion, this systematic review opens ways to academic researchers to make other data extractions narrow scope on ML types previously obtained by this research.

References

Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning (Vol. 4, No. 4, p. 122-1128). New York: springer.

Covidence systematic review software (2024). Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org

Ferlay J, Ervik M, Lam F, Laversanne M, Colombet M, Mery L, Piñeros M, Znaor A, Soerjomataram I, Bray F (2024). Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer. Available from: https://gco.iarc.who.int/today, accessed [23 Feb 2025].

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. International journal of cancer, 149(4), 778-789.

Harrison, J. E., Weber, S., Jakob, R., & Chute, C. G. (2021). ICD-11: an international classification of diseases for the twenty-first century. BMC medical informatics and decision making, 21, 1-10.

Higgins Jpt, et al. (2024). Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024). Cochrane, 2024. Available from www.training.cochrane.org/handbook.

Islam, R., Sultana, A., & Islam, M. R. (2024). A comprehensive review for chronic disease prediction using machine learning algorithms. Journal of Electrical Systems and Information Technology, 11(1), 27.

Kourou, K., Exarchos, K. P., Papaloukas, C., Sakaloglou, P., Exarchos, T., & Fotiadis, D. I. (2021). Applied machine learning in cancer research: A systematic review for patient diagnosis, classification and prognosis. Computational and Structural Biotechnology Journal, 19, 5546-5555.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews, 4(1).

Quer, G., Arnaout, R., Henne, M., & Arnaout, R. (2021). Machine learning and the future of cardiovascular care: JACC state-of-the-art review. Journal of the American College of Cardiology, 77(3), 300-313.

Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach (4th ed.). Pearson.

Saxena, A. (2021). Artificial intelligence and machine learning in healthcare. S. Chandra (Ed.). Singapore: Springer.

Siddiqui, M. F., Reza, A. W., & Kanesan, J. (2015). An automated and intelligent medical decision support system for brain MRI scans classification. PloS one, 10(8), e0135875.

Topol, E. (2019). Deep medicine: how artificial intelligence can make healthcare human again. Hachette UK.

Triantafyllidis, A. K., & Tsanas, A. (2019). Applications of machine learning in real-life digital health interventions: review of the literature. Journal of medical Internet research, 21(4), e12286.

World Health Organization. (n.d.). International Classification of Diseases (ICD). Retrieved from https://www.who.int/classifications/icd/en/

Appendix I

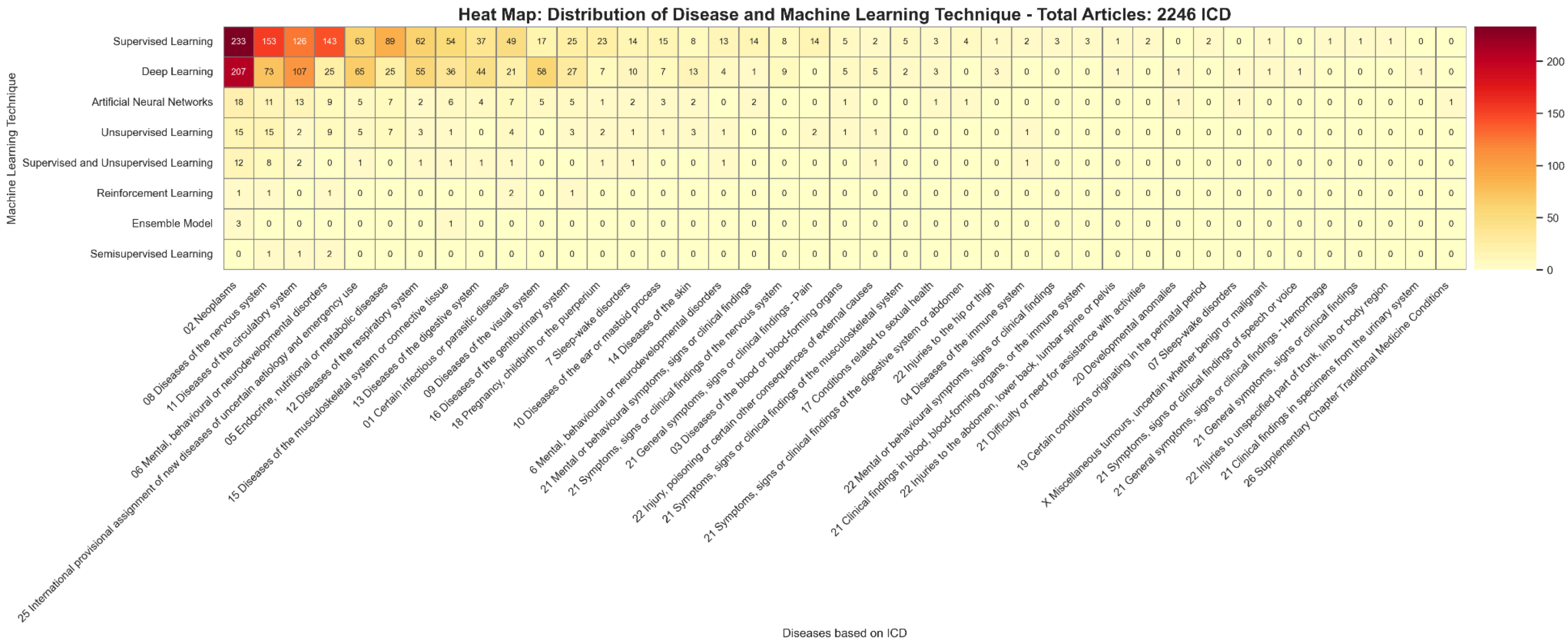

Figure 3. Heatmap representing the distribution of articles by Machine Learning technique and disease according ICD-11.

Appendix II

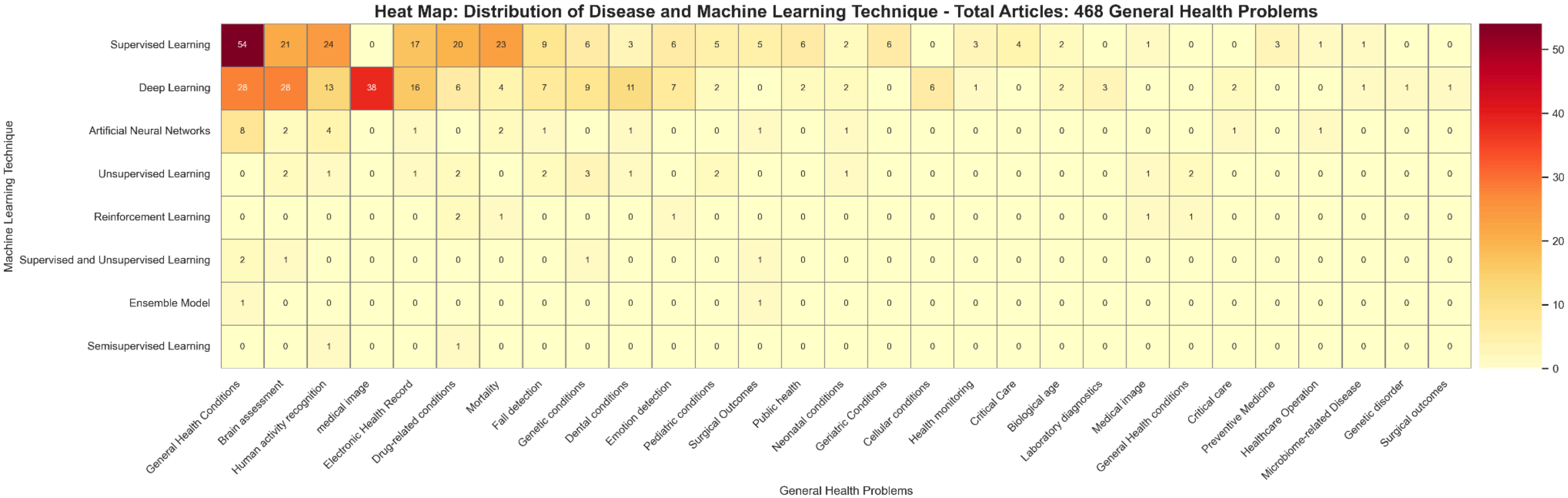

Figure 4. Heatmap representing the distribution of articles by Machine Learning technique and general health problems.