uc3m | Universidad Carlos III de Madrid

Master's Degree in Computer Science and Technology
2022-2023

Master's Thesis

Continually Adaptive Machine Learning applied to a Healthcare Platform for the Detection of Tuberculosis

Simón E. Sánchez Viloria

Jesús Carretero PhD.
Lara Visuña Perez

Leganés, September 2023

ABSTRACT

Keywords: tuberculosis, machine learning, data analysis, era4tb, healthcare

DEDICATION

Special thanks to my advisor, for guiding me and supporting me at every step in the making of this thesis . . .

> “You must know no one rejects, dislikes, or avoids pain because it is pain, but because occasionally circumstances occur in which toil and pain can procure great pleasure [. . .] In a time of freedom, when our power of choice is untrammeled and when nothing prevents us from doing what we like best, every pleasure is to be welcomed and every pain avoided [. . .] But the wise man should always hold himself to the following principle of selection: *Reject pleasure to secure greater pleasures, or else endure pains to avoid worse pains.*”
>
> *- Marcus Tullius Cicero, 45 BC*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

As health information becomes increasingly digitized, machine learning (ML) algorithms play a crucial role in deriving meaningful insights from complex, multi-modal data that is often difficult to interpret by humans. Conventional machine-learning approaches, however, may be inadequate in the face of rapidly evolving health technology, shifting patient needs, and the increasing availability of large amounts of uncertain and noisy data that is deemed too unreliable for use in clinical settings.

Modern techniques aim to address these limitations by refining models in a continual loop, prioritizing data acquisition, labeling, and feedback from experts to dynamically adapt to the data stream reliably. This work explores the potential of these techniques in the context of healthcare applications.

In particular, we incorporate the use of continual and active learning methods in the design of an end-to-end platform for the diagnosis of *tuberculosis* using computer vision techniques. Tuberculosis (TB) is an infectious pulmonary disease that has affected humankind for well over 4,000 years [1] and still affects millions of people worldwide. According to the World Health Organization, until the arrival of the COVID-19 pandemic, TB was the leading cause of widespread death from a single infectious agent, ranking even above HIV/AIDS [2].

The platform is meant to be integrated into a Healthcare and Data Portal that is being developed as part of an ongoing European project for the research of Tuberculosis, and it will serve as a tool to test and validate the use of machine-learning models for the diagnosis of Tuberculosis.

As such, in an effort to research the competence of continual adaptation methods in that context, the platform incorporates a system that is designed to continually adapt to new data as it becomes available and to ease labeling efforts by prioritizing the acquisition of the most informative data samples. Including a front-end interface for the visualization of the detection, labeling, and interaction with the models.

We evaluate the system's performance on a real dataset and compare it to a baseline model that doesn't use adaptation techniques. Our results show that the use of these methods outperforms the baseline in terms of robustness and sample efficiency while maintaining a similar level of accuracy on the test set. Furthermore, the system proposed can automatically adapt to new data and improve its performance over time.

These results demonstrate the potential of incorporating these techniques into designing real-world systems for healthcare applications and other high-stake domains. The final chapter of this work includes a lengthy discussion about possible future work and improvements to the system that might facilitate its integration into other projects and applications.

1.1. Context and Motivation

This work is done in the context of the European Regimen Accelerator for Tuberculosis (ERA4TB). ERA4TB is a public-private initiative that started in 2020 and aims to create an open European platform to accelerate the development of new treatment regimens for tuberculosis (TB). The project is integrated by over 31 organizations from the European Union and the United States, including academic institutions, research centers, non-profit organizations, and other public and private entities [3].

ERA4TB's mission statement aligns itself with (and is in response to) the United Nation's (UN) Sustainable Development Goals (SGD). The project's website reads, 'The goal of ERA4TB is to deliver an innovative and differentiated combination regimen for the treatment of TB, which can play a key role in the TB elimination agenda' [3]. This aligns with the UN's SGD target to end the TB epidemic by 2030 [4].

To address some of the challenges of TB drug development and clinical trial design, one of the project's objectives from deliverable 1.15 of ERA4TB's agenda is developing a data-science-specific platform to enable the efficient use of machine learning methods from the collaborative platform. The platform is meant to be used by researchers and other project stakeholders to facilitate the use, development, and evaluation of machine-learning models that can aid in the research of TB.

Thus, the motivation behind this work comes from the idea of incorporating novel methods into this data-science-specific platform to support researchers and collaborators in their mission to end TB by 2030. The techniques described in this work are designed to improve the performance of supervised models while prioritizing their overall robustness and reliability, aspects that are crucial in the context of healthcare applications.

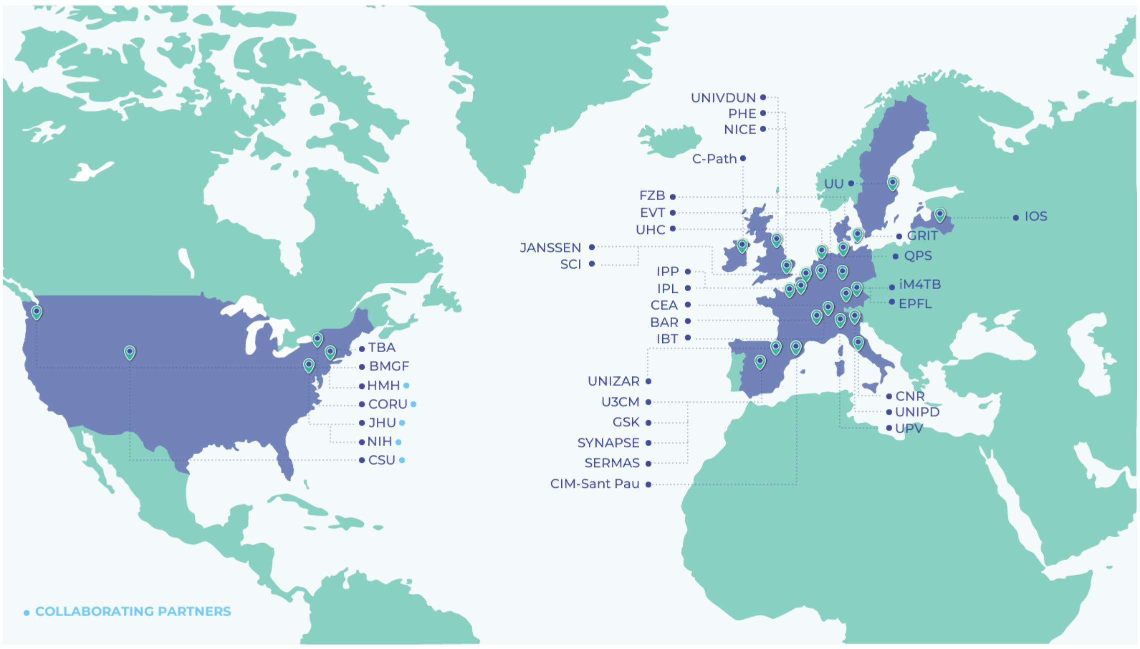

Fig. 1.1. ERA4TB's Consortium of Partners and Collaborators. Source: [3]

1.2. Background Concepts

The following section provides a brief introduction to some of the concepts and techniques that are relevant to this work. It is meant to provide the reader with the necessary high-level information to understand the context and motivation behind the ideas proposed that were used to inform every design decision and experiment conducted. For a more in-depth overview of the same concepts, the reader is referred to the literature review in Chapter 2.

1.2.1. Tuberculosis Treatment and Diagnosis

Tuberculosis is caused by the bacillus (bacteria) *Mycobacterium tuberculosis* (MTB), which is transmitted when people who are sick with TB expel the bacteria into the air by coughing, sneezing, or spitting. The disease is preventable with the administration of a vaccine and curable with the use of antibiotics over a significant period (although drug-resistant strains of the bacteria are becoming increasingly common) [5].

Nonetheless, the disease is often underdiagnosed and undertreated, especially in low-resource settings (i.e., developing countries, rural areas, and marginalized/impoverished communities), where the disease is more prevalent, calling for the development of more efficient and cost-effective methodologies to diagnose and treat the disease [2], [5], [6].

Some of the diagnosis techniques to detect TB include chest X-rays, sputum smear microscopy, and molecular tests. *Chest X-rays* are a common procedure to diagnose any signs of tuberculosis due to the wide availability of radiology devices, but they are often inconclusive and require expert radiologists to interpret the results [7].

Sputum smear microscopy is another widely available technique that requires a trained clinician to identify the bacteria under images taken with a microscope of a patient's sputum (a mixture of saliva and mucus from the respiratory tract). This technique is relatively inexpensive but requires a high concentration of bacteria in the sample to be effective. Additionally, studies argue that in conditions with limited resources and a significant number of samples, there have been reports of poor sample observation and quality control measures, which can result in false-negative results [8].

Molecular tests based on *nucleic acid amplification* (NAATs), similar to the ones popularized to detect COVID-19, are another technique to diagnose TB. These tests identify the presence of bacilli by amplifying the genetic material of the bacteria in a patient's sputum sample (if any is present) and using a chemical solution to react to it [9]. NAATs are by far the most reliable method to diagnose TB and have the advantage of being rapid and fully automated.

However, NAATs are expensive to produce and require specialized equipment, making them less accessible in low-resource settings [10], [11]. Indeed, a study conducted in 2018 showed that the ratio of smear microscopy tests to NAATs in countries with a high burden of TB was 6:1 [11], [12].

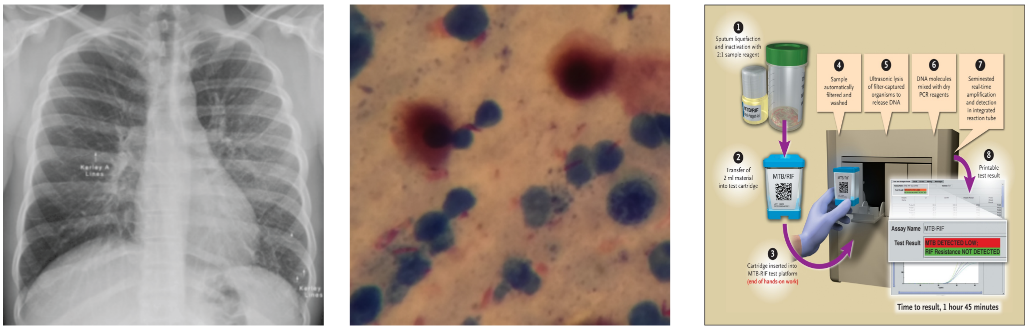

Fig. 1.2. Examples of common techniques to diagnose Tuberculosis. Left: Chest X-ray of a patient with TB [13]. Middle: Sputum smear with tuberculosis bacilli [14]. Right: Example of a molecular test for MTB [15].

Each of these techniques presents its own benefits and limitations. Recently, there have been efforts to develop machine-learning models that can aid in the diagnosis of TB as a way to reduce the need (or provide a first/second opinion) for expert clinicians in the process and improve the speed and/or accuracy of the diagnosis [1].

1.2.2. Supervised and Semi-Supervised Machine-Learning

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that studies the design of algorithms that can learn from data and make predictions based on it. ML algorithms are fed a set of data samples (often called the training set) and learn a function that maps the input data to a desired output. The goal is to learn a function that can generalize well to unseen data and make accurate predictions.

Supervised learning is a paradigm of ML where the goal is to fit a function $f : x \to \hat{y}$ that maps a given input, x, to a 'prediction' output, $\hat{y}$, based on an available finite set of input-output pairs (x_i, y_i) that are passed to a learnable model as 'training' data. The function is learned by minimizing a loss function $L(y, \hat{y})$ that measures the difference between the 'predicted' output and the actual one and returns a value that represents the error of the prediction, which is then used to update the model's parameters.

The most common supervised learning tasks are classification and regression, where the goal is to predict discrete and continuous outputs, respectively. Conversely, the aforementioned loss function is often defined based on the task at hand and the type of data available (e.g., cross-entropy loss for classification tasks, mean squared error for regression tasks, etc.).

The presence of a known output (or label) for each input is what makes this paradigm 'supervised'. For example, in healthcare, supervised learning might be used to predict the presence of a disease or condition based on a set of features extracted (or learned) from the patient's data. A model can be trained to predict the presence or severity of a disease based on a set of symptoms, clinical history, or imaging data.

[1] A literature review on the topic of TB diagnosis using ML methods is presented in section 2.2.2

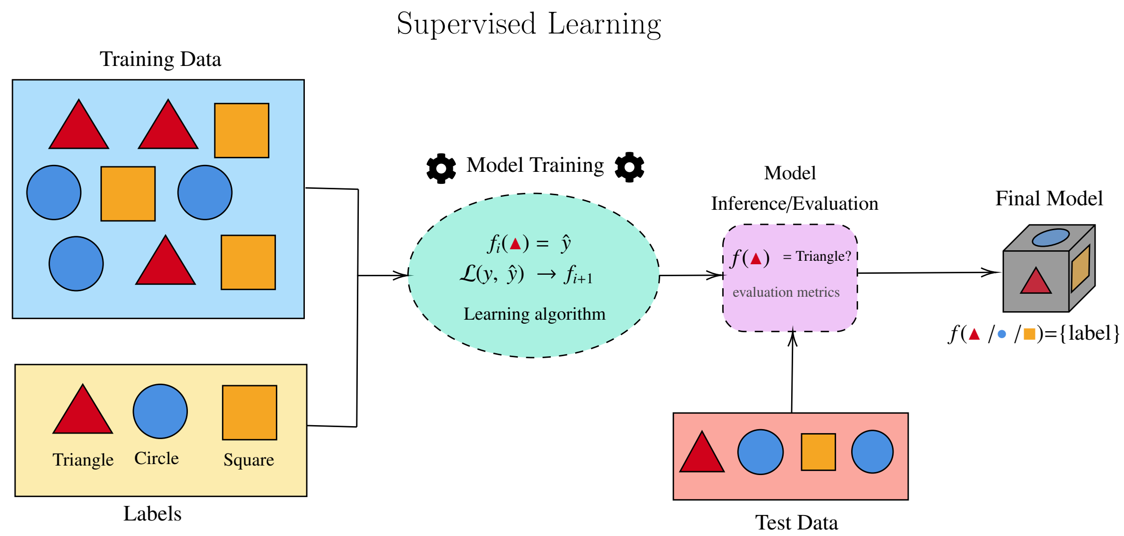

Fig. 1.3. Illustrative example of a supervised machine-learning pipeline.

Besides its many benefits, the most significant disadvantage of supervised learning is that it often needs large amounts of labeled samples to produce accurate and robust results [16], approaches such as *semi-supervised learning* (sometimes also called weak supervision) aim to address this issue by leveraging both labeled and unlabeled data to train the model [17].

Semi-supervised learning works by using unlabeled samples to learn an intermediary representation of the data that can be used as a first step to train a supervised model. This approach is especially useful when the unlabeled data is abundant and easy to obtain, but their labels are scarce and expensive. This is often the case in some healthcare applications, where the labeling task can only be performed by professional workers, making it a costly and time-consuming task [18]–[20].

In the two decades, one set of algorithms that have enabled significant advances in Machine Learning, allowing to solve very difficult problems, is **deep learning**. Deep learning is a subfield of ML that studies the design of algorithms that can learn complex representations of data by hierarchically composing simpler functions in an architecture inspired by the structure of the human brain known as ‘Deep Neural Networks’ (DNN) [16]. Figure 1.4 shows an example of a DNN architecture.

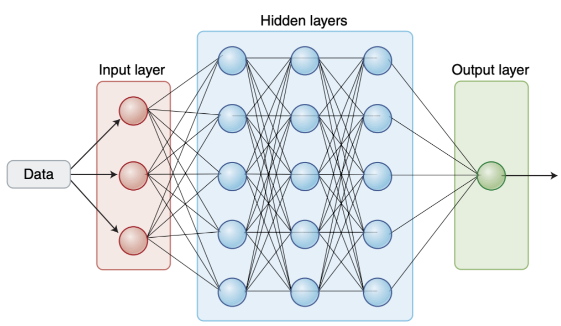

Fig. 1.4. Example of a simplified neural network architecture taken from [21].

1.2.3. Challenges and Potential of Adopting ML-enabled Systems in Healthcare

ML techniques have seen their adoption in many applications, from malware and spam detection to self-driving cars and environmental modeling. The healthcare field is no exception. In the 20 years between 1995 and 2015, the FDA had approved fewer than 30 algorithms for medical use. In contrast, only in the last 5 years, the number of new approvals has reached over *10 times* that amount (see Figure 1.5).

Indeed, the recent availability to store and process ever larger amounts of data through the use of Big Data technologies and the development of more powerful hardware and algorithmic techniques have made it possible to train models that can perform complex medical tasks with enough high accuracy and robustness to be considered for use [21].

Of such techniques, some that have recently disrupted the medical field are those based on **computer vision** (CV). CV methods try to develop algorithms that enable computers to solve visual tasks. It is a broad field that has been applied to problems like image classification, object detection, and image segmentation and has seen significant advances in the last decade thanks to the adoption of Deep Learning algorithms [16].

In the context of healthcare, CV techniques have been primarily used in radiology to aid with the diagnosis of diseases and other tasks through the analysis of medical images (e.g., X-rays, CT scans, MRIs, microscopy, etc.) [22]. Indeed, over 75% of all FDA-cleared AI applications have been for radiology use cases [23].

One study made in 2018 trained a *Convolutional Neural Network* (CNN), a common DL model popular for computer vision tasks, to detect pneumonia from X-ray images at an Indian hospital. The researchers compared the performance of the algorithm with the findings of four expert radiologists and concluded that the algorithm was comparable to (sometimes even *outperformed*) the radiologist in most cases [24].

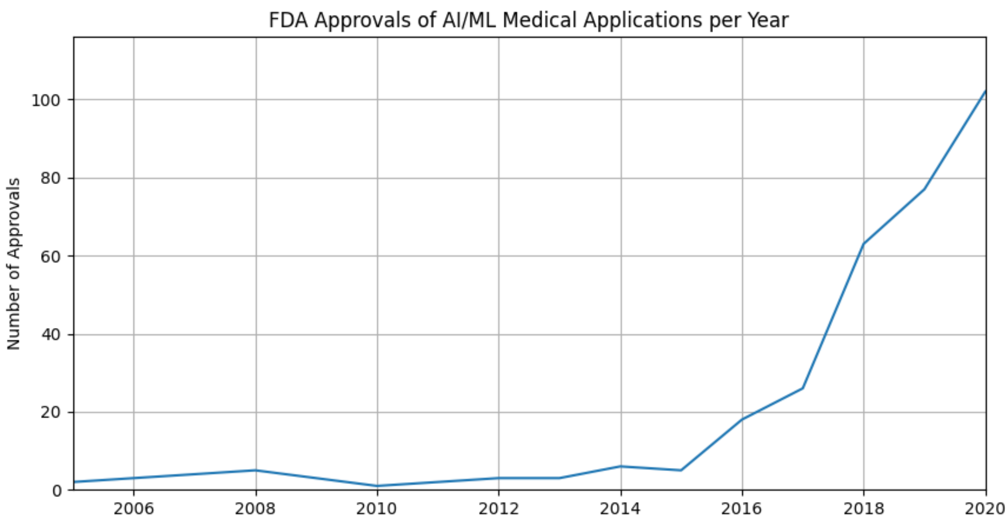

Fig. 1.5. Number of FDA-approved AI applications per year since 2005. Data Source: [23]

Other AI systems such as Google Deepmind's *AlphaFold* [25] have been shown to predict the 3D structure of proteins with high accuracy, solving a problem that had been considered to be one of the most challenging in computational biology for over 50 years. Such breakthrough is thought to have a significant impact in applications like drug discovery in the near future [26].

Results such as these shine a light on the potential that AI techniques have in the medical field. However, the adoption of such technologies doesn't come with new challenges. Experts have emphasized the importance of improving aspects of these models like their lack of interpretability, robustness to unseen data, and the difficulty of integrating them into existing workflows before they can be widely adopted in clinical settings [21], [22].

Thus, the design of ml-enabled systems must be mindful of the limitations of such models. Research in the healthcare and AI fields must pave the way to develop workflows that can be trusted by clinicians, patients, lawmakers, and other researchers alike to improve the quality of care/research, reduce costs, and overall increase the efficiency of the healthcare system.

Company	Year	Usecase	Panel
Apple	2022	Atrial Fibrillation Detection via Apple Watch	Cardiovascular
Arterys	2022	Liver and Lung Cancer Detection	Radiology
Philips Healthcare	2022	Philips Incisive CT Reconstruction	Radiology
GE Healthcare	2021	Deep Learning Image Reconstruction	Radiology
Siemens	2021	AI-Rad Companion for CT Interpretation	Radiology
Icometrix	2018	Brain MRI Analysis	Radiology
23&Me	2017	Genetic Testing for Hereditary Thrombophilia	Hematology

Table 1.1. EXAMPLES OF FDA-APPROVED AI APPLICATIONS FOR MEDICAL USE. Source: [23]

1.2.4. Self-Adaptive Systems

Table 1.2 lists some of the most common causes of degradation of a learning system and the techniques that can be used to address them. The causes and descriptions in the table were sourced from Casmiro et al.'s work on self-adaptive machine-learning [27] and Chip Huyen's book about Machine-Learning systems [28] [27].

Add a brief description of the concept of continual adaptation and its relevance in the context of this work.

Cause	Description	Example
Data distribution shift	**Covariate shift:** When the input distribution $P(X)$ that the model was trained with differs significantly from the one in the inference environment (i.e. the input changes over time, but the model remains the same).	A model trained on chest X-ray images from a particular dataset is deployed in a hospital where the X-ray machines are of a different brand and produce images with different characteristics.
	Label shift: When the distribution $P(Y)$ at inference time differs from the one the model was trained with, i.e. the model is trained on a dataset whose class proportions are substantially different from the ones in the inference environment.	A model trained on a dataset where the proportion of positive cases is 50% is deployed in a hospital where the proportion is 10%.
	Concept shift: The relationship between the input and output changes over time. That is, the model is trained on a dataset whose relationship between the input and output differs from the one in the inference environment.	
Model	**Model drift:**The model's performance degrades over time.	?
	Model bias: The model is biased towards a specific class or group.	
Environment drift	The environment changes over time. The model is deployed in a different environment than the one it was trained on.	

Table 1.2. SOME CAUSES OF DEGRADATION OF AN ML SYSTEM

1.3. Objectives

Main objective

The main objective of this work is to research the most relevant techniques on continual adaptation methods in Machine Learning, make a comparative analysis of the most relevant techniques and their relevance for health applications, and design and implement a system for the diagnosis of tuberculosis that incorporates these techniques into its design that can be integrated into the ERA4TB platform. The platform should allow its users to incorporate test machine-learning models into the platform and facilitate the use and collaboration between researchers and other stakeholders.

Specific objectives

Auxiliary to the main objective, the following specific objectives have been defined to guide the development of the work and evaluate its success:

1. The system should be capable of automatically triggering the continual learning process when new data is available or when the model's performance degrades based on a predefined metric.
2. The platform must implement a feedback loop between the data annotation process and model training that prioritizes the acquisition of the most informative data samples to improve the model's performance (Active Learning).
3. Develop a front-end interface that allows users to interact with the machine-learning models by selecting or submitting new data samples and visualizing the model's predictions.
4. Consider the limitations of the proposed system and the ethical implications of its use and present a well-founded outline of necessary future work to address these limitations or improve the system in a way that aligns with the project's mission statement.
5. Evaluate possible future research directions that could be explored in the area of continual and dynamic adaptation in Machine Learning, highlighting the contributions that have a higher potential for impact in healthcare or other high-stake domains.

1.4. Main Contributions

1.5. Structure of the Work

This work is divided into five chapters, including this introduction. Chapter 1 describes the context and motivation of this work and its objectives and provides the necessary background information to understand the concepts and techniques used.

Chapter 2 describes relevant work and state-of-the-art techniques. It also provides a literature review of related work in tuberculosis detection and adaptive machine learning systems, highlighting the most important contributions and their limitations and showing examples of their use in healthcare applications.

Chapter 3 describes the methodology used to design and implement the proposed system. It gives a detailed description of the data, models, techniques, and specific tools used in the implementation of the platform, experiments conducted, and relevant metrics to evaluate the system's performance.

Chapter 4 presents the results of the experiments described in the previous chapter, analyzing the performance of the system, and comparing the results with the baseline metrics.

Finally, Chapter 5 presents the conclusions of the work. It discusses the implications of the results obtained in the context of the project and the limitations of the proposed system. It also proposes possible future work that could be explored to improve the system and its integration into the ERA4TB platform and discusses emergent directions in the areas of research discussed in this work, highlighting those that have a higher potential for impact or that we consider to be of special interest.

Additionally, there are two extra chapters at the end where we discuss the regulatory framework, ethical considerations, and the socioeconomic implications of the use of the proposed system (or similar systems) in the context of the project and the healthcare field in general.

2. STATE OF THE ART

10-12 pgs

2.1. Relevant Techniques

6-7 pages

While it is important to consider the design of robust machine learning models from their training phase, and part of this work will discuss techniques to optimize for that goal, our primary concern is to provide a model-agnostic framework for maintaining reliability *after* a model has been already deployed.

To accomplish this, we consider a set of novel techniques proposed in the literature to address the problems related to deploying supervised and semi-supervised machine learning models in real-world scenarios. The following section describes these techniques and their relevance to the problem at hand.

Note that while the focus is on applying these techniques to healthcare applications, it is easy to show that these methods can be generalized to other domains.

2.1.1. Transfer Learning and Domain Adaptation

1 page

Transfer learning is a technique that aims to improve the performance of a model by transferring the knowledge of a pre-trained model to a new task. This is achieved by training the model on a large dataset and then fine-tuning it on a smaller dataset for the new task. The model is trained on a large dataset to learn general features that can be applied to a wide range of tasks. The model is then 'fine-tuned' on the smaller dataset to learn task-specific features that are relevant to the new task **pan_survey_2010**.

2.1.2. Continual Learning

1.25 pages

Continual learning refers to the concept of constantly updating a model as new information arrives, allowing them to adapt to changing data [28]. This constant change in the data distribution requires models to be updated periodically to maintain their accuracy and relevance, prevent model stagnation and extend their relevance to the application.

We can consider two frameworks for implementing continual learning in a machine-learning system. We refer to the first as *offline learning*, where the model is updated periodically using a batch of data collected over time. Offline learning can be additionally subdivided into two categories based on how data is monitored prior to updating a model: *passive* and *active* continual learning.

The model is updated passively when the data is collected over a fixed period of time (e.g., every six months) or after a fixed number of instances have been processed and labeled (e.g., every 1000 new instances). Conversely, the model is updated actively when we update only when the model's performance drops below a certain threshold or when the system detects a significant change in the data distribution [28].

Online learning is the second framework for implementing continual learning. The difference between online and offline learning is that, with the former, the model is updated as soon as new data arrives, with every new instance - or small batch of instances - being used to update the model. Because this approach tends to be more computationally expensive than offline learning and suffers from well-known problems such as catastrophic forgetting [28], it is often used in conjunction with offline learning to improve the model's performance.

Continual learning is well-suited for healthcare applications where data arrives in a stream (e.g., wearable health monitors, electronic medical records, and imaging systems). It is also useful in terms of

2.1.3. Active Learning

1.25 pages

Active learning strategies selectively acquire data based on their informativeness or uncertainty to the model. Its value comes from allowing the model to guide its own data acquisition process, thus potentially reducing the need for vast - or unnecessary - amounts of pre-labeled data before a model is trained or updated [19], [20], [28]. This is particularly important in health applications, as the availability of annotated medical data is often limited due to privacy concerns, expert time constraints, and the complexity of pathological findings.

Active learning enables the development of accurate models using significantly less labeled data, paving the way for more efficient and cost-effective machine-learning platforms. Some of the most common active learning strategies include uncertainty sampling, query by committee, and expected model change.

2.1.4. Knowledge Distillation

0.5 pages

Knowledge distillation is a technique that aims to improve the performance of a model by transferring the knowledge of a larger model (teacher) to a smaller model (student). This is achieved by training the student model to mimic the predictions of the teacher model. The student model is trained on the same data as the teacher model, but it is trained to predict the probabilities of the teacher model's predictions instead of the actual labels. This allows the student model to learn from the teacher model's mistakes and improve its performance on the given task [29].

Modify description, add more details, references ...

2.1.5. Dynamic Quantization and Network Pruning

0.5 pages

Quantization and network pruning are techniques that aim to reduce the size of a model by removing redundant parameters...

Cite [30] and [31], [32]

Modify this description and add more details, references ...

2.1.6. Federated Learning

1 page

Federated learning is the concept of training a model using data from multiple sources without having to share the data itself. This is achieved by training the model on each source separately and then combining the results to obtain a final model. This technique is particularly useful in healthcare applications, where data privacy is a major concern. It allows us to train models on data from multiple sources without having to share the data itself, thus preserving the privacy of the patients [33].

Modify this description and add more details, references ...

Add a full-page figure detailing the different techniques. You can use Matcha (https://matcha.io/) to create a pretty image and export it as a tikz figure.

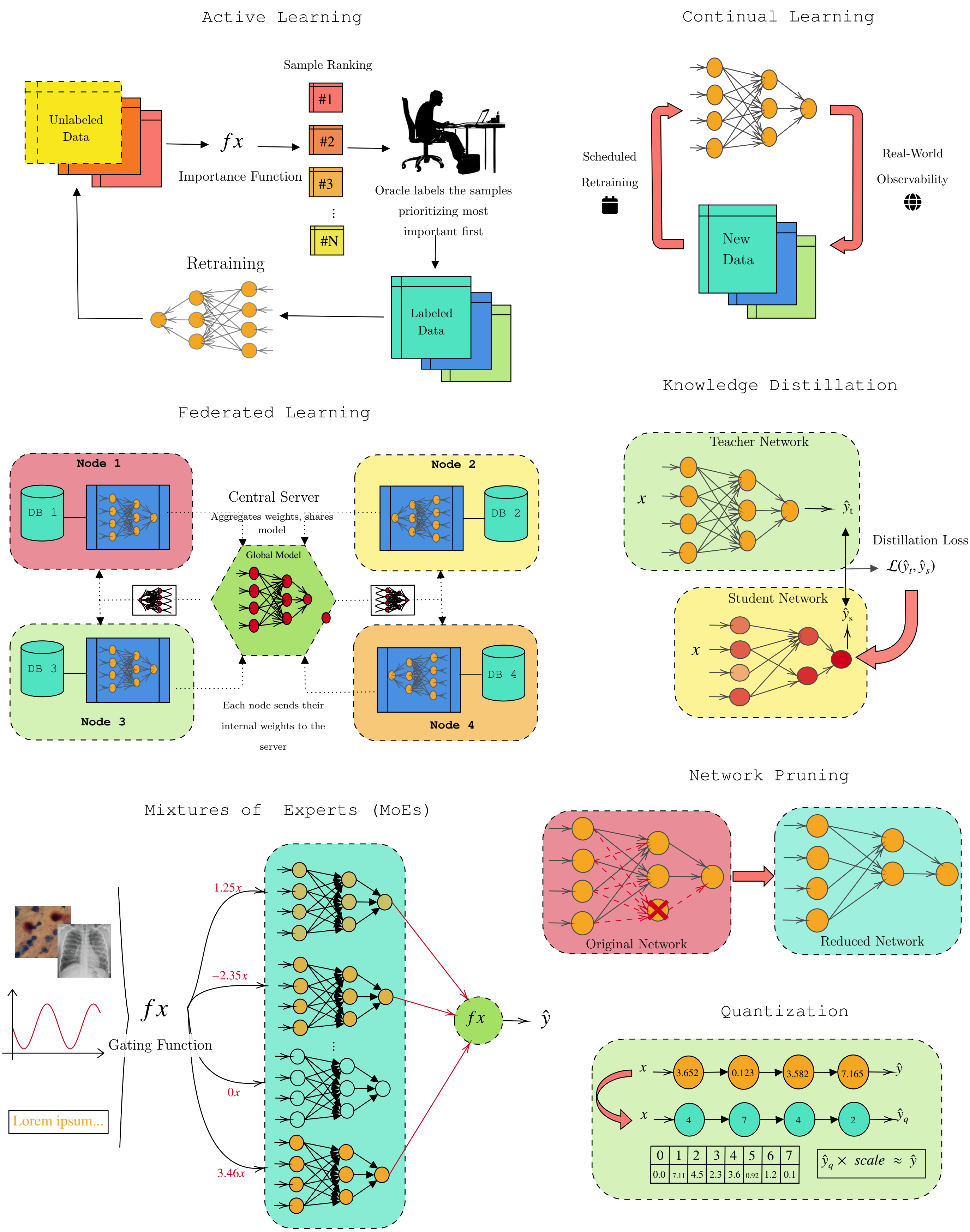

Fig. 2.1. Relevant Machine Learning Paradigms and Techniques

2.2. Related Work

2.2.1. Computer Vision in Healthcare

2 pages

SOTA methods in computer vision with an emphasis on object detection techniques (CNN, YOLO, Faster R-CNN, etc.)

2.2.2. Tuberculosis Detection using Machine Learning Methods

1-1.5 pages

Visuña et al. (2023) [34] presented a deep-learning-based technique to detect tuberculosis from sputum smear microscopy images. The author used a one-stage object detection method with a Convolutional Neural Network (CNN) backbone to detect the presence of bacilli in the images, first fragmenting the image into patches of 80x80 pixels. The model was trained on a dataset of 200 microscopy stain images, achieving a 99.49% precision and 92.86% recall on the test set.

2.2.3. Continually Adaptive Systems

1-1.5 pages

Include a full-page table with a summary of each reference and their exact contribution to this work, use different columns to highlight the differences between them and the techniques used. Only include those whose techniques are implemented in the proposed solution (continual, active learning, tuberculosis detection, adaptive systems, etc.). If you are unsure about how well the table fits here, perhaps consider moving it to the appendix.

Reference	Domain	Technique(s)	Summary of Contributions
Visuña et al. (2023) [34]	Tuberculosis detection	Object detection, Image pre-processing, CNN fine-tuning, NASNetMobile	Tested novel DL-based method on 200 microscopy stain images of sputum. Reached 99.49% precision and 92.86% recall

Table 2.1. SUMMARY OF THE MOST RELEVANT WORKS IN THE LITERATURE.

3. DESIGN OF THE SOLUTION

This work is driven by the problem of designing platforms primarily used to perform Machine-Learning inference (only obtaining the output predictions from a model) while ensuring that the obtained outputs are as reliable and robust as possible when encountering new data. This is important in scenarios where a model is deployed in high-stake environments where their output causes significant downstream impact, and inaccurate predictions may be costly to the relevant stakeholders.

We consider that these types of platforms have the following characteristics:

- The (already trained) models are uploaded to the platform primarily to obtain predictions from new data. We refer to the bag of models available to the platform as the *model repository*.
- The data used to train and evaluate the models is available. This means the platform can have prior knowledge about the data distribution the models were trained on.
- The platform has enough hardware resources available to perform inference with the models that it has available.
- The platform constantly receives new data for inference, which is assumed to be independent of the data used to train the models.
- Labeling capabilities are limited. That is, annotating new data is costly and unfeasible in large amounts. This is common in real-world applications, notably in the healthcare sector, where the labeling task is performed by professional workers, making it a costly and time-consuming task [18]–[20].

Furthermore, for the purpose of this work, we consider that the inference environment is deployed alongside an *experimental* environment. This environment is assumed to be detached from the training environment and is set up to continuously evaluate the models' robustness on new data, evaluate its performance after retraining using the aforementioned techniques, and compare the results with the deployed model. If the experimental model outperforms the deployed model, the latter is updated with the new model. This process is repeated for the entire lifetime of the application.

The role of the human annotator in designing machine-learning models is also an important aspect of this work...

4. RESULTS

In this chapter, we present a rundown of the results obtained from evaluating all the experiments. First, an analysis of the results of testing our system...

4.1. Analysis of the results

.

4.2. Comparison of the different methods

4.3. Discussion

5. CONCLUSIONS

9-10 páginas

Discussion about the results obtained and implications in the context of the project, limitations of the proposed system, future work, etc.

5.1. Main Implications

0.65 pág

5.2. Limitations of the System

0.5 pág

5.3. Future Work

2.75 pág

The following section...

5.3.1. Implementing new adaptation strategies to the system

0.5 pág

5.3.2. Improving the explainability of the continual learning process

0.5 pág

Anytime a model is trained on a particular dataset, any biases present on that dataset are also introduced to the model. In any high-stakes environment, the researchers who design these models should be aware of the biases present in the data and how they might affect the model's predictions and consider them in the decision-making process.

Furthermore, this aspect is especially crucial in the healthcare sector, where legislation and ethical guidelines stress the importance of AI systems being transparent [35], [36]. Healthcare professionals have the moral and legal obligation to be able to explain the decisions made by the AI systems they use to the relevant stakeholders, the latter also has the right to know how the decisions that affect them are made.

In that regard, one of the main limitations of the system proposed here is the lack of explainability of the continual learning process, where a model that is designed initially to be as transparent as possible might gradually lose explainability power as it adapts to new data distributions over time.

Methods such as TRAK [37] have been proposed as a way to improve the explainability of Deep Learning models by providing a way to trace the predictions of a model to individual instances of the training data in a concept known as *data attribution*, which has been proven useful in improving sample selection for active learning approaches [37]–[39].

5.3.3. Adapting the System to run on a Federated Platform

1.75 pág

Efforts such as [33] have demonstrated the benefits and potential of taking a federated learning approach to training machine learning models in the healthcare industry. A future proposal to improve the system presented in this work would be to adapt the current platform to one where the models are trained in a federated fashion.

We would propose a system where each research laboratory trains an instance of a (previously agreed) global model with its own local (private) data and shares the trained weights with a central server. The new platform would be mainly used to perform inference on public data, while each client could have an instance of the annotation frontend and backend, including the active learning framework, running locally to annotate their data.

The motivation for this comes from the fact that is very difficult to convince partners (even within the same project organization) to share their data. In the healthcare industry, data is considered to be a very valuable asset, with a high cost to obtain and annotate, which is why it is often used as a bargaining chip in negotiations between companies and research organizations . include citation

Show a diagram of the proposed federated learning system.

5.4. Further Research Directions

4.5 pág

The following are some of the recent and upcoming research directions in the field of machine learning that we consider to be relevant to the work presented in this thesis.

Unlike, the previous section, which focused on the limitations of the proposed system and detailed ideas about how to address or improve them in the immediate future, this section takes a broader view, focusing on the limitations of the current state-of-the-art and what might be considered to be the next steps in the area.

The idea is that these research directions could used as a starting point for future work (maybe even a Ph.D. thesis) on the topics of adaptive machine-learning systems, either as a continuation of some of the work presented here or as a completely new approach to the problem. We make no claims about the feasibility of these ideas but rather present why I consider them to be relevant and interesting for further research and discussion.

5.4.1. Future directions in Tuberculosis AI Research

<1 pág

Detection will soon be a solved problem thanks to NAATs (or rather, more of a money problem than a technical one), thus, the emphasis should now be on drug discovery, finding new biomarkers, etc. . . .

5.4.2. Meta-Learning and Self-Improving Systems

1 pág

Much like human learners, who, building from previous knowledge, continuously seek and filter information that could be useful to learn new concepts and skills, an area of research in machine learning concerns the design of programs/systems that can efficiently improve their learning process without the need for explicit human intervention. This area of research is known as **meta-learning**, and it is a very active area of research in AI .

add citation

Meta-learning is a technique that aims to improve the performance of machine-learning models by 'learning to learn' (L2L) a certain task. Such ideas have been successfully applied to a wide range of problems, including computer vision, natural language processing, robotics, video games, and more [40].

The way meta-learning is formulated is by training a model on a variety of tasks and then using the knowledge gained from those to improve its performance on new tasks or learn it faster / more sample-efficiently than if it had been trained only for that task [40].

This idea is regarded to have been first introduced by Dr. Jurgen Schmidhuber in 1987 with his thesis 'Evolutionary Principles in Self-Referential Learning'. In his work, Schmidhuber proposed an algorithm that adaptively improves its learning skills by recursively applying genetic programming to itself and ensuring that only 'useful' modifications (made by the program to itself) 'survive' in an evolutionary fashion [41].

Recently, Finn et al. (2017) [42] propose a model-agnostic framework for meta-learning that can be applied to any deep-learning architecture and learning task. The framework consists of . . .

The idea behind researching this area is that, by adopting this L2L framework, we could design a system that can continually improve its learning process over time.

As a low-hanging fruit, we can envision the design of a system similar to the one proposed in this work that integrates and builds upon concepts from meta-learning, unlearning, knowledge distillation, and transfer learning [2], that enables model that can more sophisticatedly adapt to new problems.

This self-adaptive process would necessarily be based on the evaluation of the model's performance and a metric of the 'necessity' of adaptation/learning that task better, but rather than relying on simple heuristics and a model-agnostic approach, the system would trigger a more complex adaptation process.

[2] See section 2.1

5.4.3. Scalable Adaptability and Multimodality through Mixtures of Experts

1.5 pág

Mixture of experts (MoEs) models are a type of ensemble model that combines the predictions of multiple models to obtain a final prediction. The difference between MoEs and other ensemble models is that the predictions of the individual 'experts' are combined using a gating function that adapts to the given data point and dynamically determines the weight of each model in the final prediction [43].

MoEs are really powerful systems. They have been shown to be able to learn complex multimodal distributions and have been used in a wide variety of applications, including object detection [44], language modeling [45], machine translation [46], and even multi-omics [47].

In the context of this work, we consider that MoEs could be used to create a more scalable and robust system that can adapt to new tasks and data distributions. The idea is that the system would be composed of a set of 'expert' models, each of which would be specialized in a particular task or data distribution. The system would then be able to adapt to new tasks by autonomously learning a new expert model or by retraining an existing one.

The main advantage of this approach is that it would allow the system to scale to a large number of tasks and data distributions, only needing to retrain the gating function continually instead of entire models.

Furthermore, each expert model can be deployed independently, allowing for a more flexible, modular system capable of being distributed among different environments and hardware resources. Systems that take advantage of MoEs to scale billion-parameter models to multiple GPUs already exist and have been deployed to production for applications as big as ChatGPT . add citation

Another advantage of this approach is the potential for more failsafe systems. By having each model deployed independently in a distributed system, one could devise a mechanism that detects when one of the expert models fails, either due to a hardware/-software error or due to a significant performance drop, and automatically replaces the gating function of the current MoE with one (previously trained) that excludes the failing model.

One could also use such a mechanism as a way to save operational costs, the system might monitor the number of instances being routed to each expert model and, if one of the models is not being used, it could be automatically shut down to save resources.

Another interesting research direction would be to explore the use of MoEs to handle multimodal data. In [45] researchers at Google Brain used MoEs to train an architecture that accepts both image and text inputs and outputs a single prediction by

5.5. Final Remarks

0.75 pág

...

Unlike many scientific advances, breakthroughs such as a true self-improving model that adapts itself continually in an online fashion will have immediate applications in every domain, from healthcare to education, to economics, to scientific discovery itself ...

REGULATORY FRAMEWORK

add url to github repo

Possible references to current or possible future legislation/regulations about the use of AI for healthcare applications, healthcare data, or other related topics (e.g., GDPR, HIPAA, EU AI Act, etc.)

Ethical Considerations

The EU published in 2019 their 'Ethic guidelines for trustworthy AI' [35], which are based on seven key requirements that AI systems should meet in order to be considered trustworthy. These requirements are human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity with regard to non-discrimination and fairness, environmental and societal well-being, and accountability.

Any AI system that is developed and deployed should meet these requirements, and this work is no exception ...

Furthermore, the more recent EU AI Act [36]...

[3] All code developed in this work can be found under the following url: https://github.com/simonsanvil/..., for any questions, concerns, or comments, please contact the author at simonsviloria@gmail.com

SOCIO-ECONOMIC ENVIRONMENT

Budget

The estimated costs of the realization of this project include those related to the human labor and material costs associated with it.

In terms of human resources, both the student author of the work and the advisor of the thesis are considered to have put hours of labor into the making of this project. We have assumed the salary of the student to be equivalent to the one of a junior engineer of about €15.00 per hour of labor, and the one of the advisor to be equivalent to that of a senior engineer of around €35.00 per hour.

An estimate of xxx hours...

	# of Hours	Salary per hour (€)	Total Salary (€)
Author	xxxx	yy	zzzz
Advisor	xxx	yy	zzzz
		Total Cost:	**€...**

Table 5.1. ESTIMATED COSTS OF HUMAN RESOURCES

In terms of non-human resource costs, the only relevant ones are those associated with the material costs of the hardware utilized throughout the work. This is because only open-source software tools (Python, R, LaTeX) were utilized.

...

Socio-Economic Impact

BIBLIOGRAPHY

[1] CDCTB, *World TB Day History*, en-us, Feb. 2023. [Online]. Available: https://www.cdc.gov/tb/worldtbday/history.htm (visited on 08/19/2023).

[2] WHO, *Global Tuberculosis Report 2022*, en, 2022. [Online]. Available: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022 (visited on 08/19/2023).

[3] *European Accelerator of Tuberculosis Regime Project*, en-US. [Online]. Available: https://era4tb.org/ (visited on 04/24/2023).

[4] World Health Organization. Regional Office for Europe, "Tuberculosis: Fact sheet on Sustainable Development Goals (SDGs): Health targets," en, World Health Organization. Regional Office for Europe, Tech. Rep. WHO/EURO:2017-2388-42143-58059, 2017, number-of-pages: 8. [Online]. Available: https://apps.who.int/iris/handle/10665/340885 (visited on 08/19/2023).

[5] WHO, *Tuberculosis (TB)*, en, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/tuberculosis (visited on 08/20/2023).

[6] IMI, *IMI Innovative Medicines Initiative | ERA4TB | European regimen accelerator for tuberculosis*, en, Jan. 2020. [Online]. Available: http://www.imi.europa.eu/projects-results/project-factsheets/era4tb (visited on 08/20/2023).

[7] P. Escalante, "Tuberculosis," en, *Annals of Internal Medicine*, vol. 150, no. 11, ITC6–1, Jun. 2009. DOI: 10.7326/0003-4819-150-11-200906020-01006. [Online]. Available: http://annals.org/article.aspx?doi=10.7326/0003-4819-150-11-200906020-01006 (visited on 08/20/2023).

[8] P. Desikan, "Sputum smear microscopy in tuberculosis: Is it still relevant?" *The Indian Journal of Medical Research*, vol. 137, no. 3, pp. 442–444, Mar. 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3705651/ (visited on 08/19/2023).

[9] C. C. for Disease Control and Prevention, Governmental, *TB Diagnostic Tool: Xpert MTB/RIF Assay Fact Sheet | TB | CDC*, en-us, Aug. 2016. [Online]. Available: https://www.cdc.gov/tb/publications/factsheets/testing/xpert_mtb-rif.htm (visited on 08/19/2023).

[10] H. Albert *et al.*, "Development, roll-out and impact of Xpert MTB/RIF for tuberculosis: What lessons have we learnt and how can we do better?" eng, *The European Respiratory Journal*, vol. 48, no. 2, pp. 516–525, Aug. 2016. DOI: 10.1183/13993003.00543-2016.

[11] E. MacLean *et al.*, "Advances in Molecular Diagnosis of Tuberculosis," *Journal of Clinical Microbiology*, vol. 58, no. 10, 10.1128/jcm.01582–19, Sep. 2020, Publisher: American Society for Microbiology. DOI: 10.1128/jcm.01582-19. [Online]. Available: https://journals.asm.org/doi/10.1128/jcm.01582-19 (visited on 08/19/2023).

[12] D. Cazabon *et al.*, "Market penetration of Xpert MTB/RIF in high tuberculosis burden countries: A trend analysis from 2014 - 2016," eng, *Gates Open Research*, vol. 2, p. 35, 2018. DOI: 10.12688/gatesopenres.12842.2.

[13] B. A. A. Ubaidi, "The Radiological Diagnosis of Pulmonary Tuberculosis (TB) in Primary Care," en-US, Mar. 2018, Publisher: clinmed journals. DOI: 10.23937/2469-5793/1510073. [Online]. Available: https://www.clinmedjournals.org/articles/jfmdp/journal-of-family-medicine-and-disease-prevention-jfmdp-4-073.php?jid=jfmdp (visited on 08/22/2023).

[14] M. I. Shah *et al.*, "Ziehl–Neelsen sputum smear microscopy image database: A resource to facilitate automated bacilli detection for tuberculosis diagnosis," *Journal of Medical Imaging*, vol. 4, no. 2, p. 027 503, Apr. 2017. DOI: 10.1117/1.JMI.4.2.027503. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5492794/ (visited on 08/22/2023).

[15] C. C. Boehme *et al.*, "Rapid Molecular Detection of Tuberculosis and Rifampin Resistance," *New England Journal of Medicine*, vol. 363, no. 11, pp. 1005–1015, Sep. 2010, Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa0907847. DOI: 10.1056/NEJMoa0907847. [Online]. Available: https://doi.org/10.1056/NEJMoa0907847 (visited on 08/22/2023).

[16] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," en, *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, Number: 7553 Publisher: Nature Publishing Group. DOI: 10.1038/nature14539. [Online]. Available: https://www.nature.com/articles/nature14539 (visited on 05/06/2022).

[17] X. Zhu, "Semi-Supervised Learning Literature Survey," *Comput Sci, University of Wisconsin-Madison*, vol. 2, Jul. 2008.

[18] A. Yakimovich, A. Beaugnon, Y. Huang, and E. Ozkirimli, "Labels in a haystack: Approaches beyond supervised learning in biomedical applications," en, *Patterns*, vol. 2, no. 12, p. 100 383, Dec. 2021. DOI: 10.1016/j.patter.2021.100383. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666389921002506 (visited on 04/24/2023).

[19] Y. Chen, T. A. Lasko, Q. Mei, J. C. Denny, and H. Xu, "A study of active learning methods for named entity recognition in clinical text," en, *Journal of Biomedical Informatics*, vol. 58, pp. 11–18, Dec. 2015. DOI: 10.1016/j.jbi.2015.09.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046415002038 (visited on 04/24/2023).

[20] R. L. Figueroa, Q. Zeng-Treitler, S. Kandula, and L. H. Ngo, “Predicting sample size required for classification performance,” *BMC Medical Informatics and Decision Making*, vol. 12, no. 1, p. 8, Feb. 2012. DOI: 10.1186/1472-6947-12-8. [Online]. Available: https://doi.org/10.1186/1472-6947-12-8 (visited on 04/24/2023).

[21] E. J. Topol, “High-performance medicine: The convergence of human and artificial intelligence,” en, *Nature Medicine*, vol. 25, no. 1, pp. 44–56, Jan. 2019, Number: 1 Publisher: Nature Publishing Group. DOI: 10.1038/s41591-018-0300-7. [Online]. Available: https://www.nature.com/articles/s41591-018-0300-7 (visited on 08/25/2023).

[22] A. Esteva *et al.*, “Deep learning-enabled medical computer vision,” *NPJ Digital Medicine*, vol. 4, p. 5, Jan. 2021. DOI: 10.1038/s41746-020-00376-2. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7794558/ (visited on 08/25/2023).

[23] C. f. D. a. R. Health, “Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices,” en, *FDA*, Oct. 2022, Publisher: FDA. [Online]. Available: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices (visited on 08/26/2023).

[24] X. Wang *et al.*, “ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, arXiv:1705.02315 [cs], Jul. 2017, pp. 3462–3471. DOI: 10.1109/CVPR.2017.369. [Online]. Available: http://arxiv.org/abs/1705.02315 (visited on 08/25/2023).

[25] Z. Yang, X. Zeng, Y. Zhao, and R. Chen, “AlphaFold2 and its applications in the fields of biology and medicine,” en, *Signal Transduction and Targeted Therapy*, vol. 8, no. 1, pp. 1–14, Mar. 2023, Number: 1 Publisher: Nature Publishing Group. DOI: 10.1038/s41392-023-01381-z. [Online]. Available: https://www.nature.com/articles/s41392-023-01381-z (visited on 08/26/2023).

[26] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” en, *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, Number: 7873 Publisher: Nature Publishing Group. DOI: 10.1038/s41586-021-03819-2. [Online]. Available: https://www.nature.com/articles/s41586-021-03819-2 (visited on 08/26/2023).

[27] M. Casimiro *et al.*, “Self-adaptive Machine Learning Systems: Research Challenges and Opportunities,” in Aug. 2022, pp. 133–155. DOI: 10.1007/978-3-031-15116-3_7.

[28] C. Huyen, *Designing Machine Learning Systems*, en. O'Reilly Media, Inc., May 2022, ISBN: 9781098107963. [Online]. Available: https://www.oreilly.com/library/view/designing-machine-learning/9781098107956/ (visited on 04/24/2023).

[29] G. Hinton, O. Vinyals, and J. Dean, *Distilling the Knowledge in a Neural Network*, arXiv:1503.02531 [cs, stat], Mar. 2015. DOI: 10.48550/arXiv.1503.02531. [Online]. Available: http://arxiv.org/abs/1503.02531 (visited on 08/20/2023).

[30] M. Á. Carreira-Perpiñán and Y. Idelbayev, *Model compression as constrained optimization, with application to neural nets. Part II: Quantization*, arXiv:1707.04319 [cs, math, stat], Jul. 2017. DOI: 10.48550/arXiv.1707.04319. [Online]. Available: http://arxiv.org/abs/1707.04319 (visited on 08/22/2023).

[31] S. Han, H. Mao, and W. J. Dally, *Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding*, arXiv:1510.00149 [cs], Feb. 2016. [Online]. Available: http://arxiv.org/abs/1510.00149 (visited on 08/22/2023).

[32] M. A. Carreira-Perpinan and Y. Idelbayev, “"Learning-Compression" Algorithms for Neural Net Pruning,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, ISSN: 2575-7075, Jun. 2018, pp. 8532–8541. DOI: 10.1109/CVPR.2018.00890.

[33] M. Joshi, A. Pal, and M. Sankarasubbu, “Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges,” en, *ACM Transactions on Computing for Healthcare*, vol. 3, no. 4, pp. 1–36, Oct. 2022. DOI: 10.1145/3533708. [Online]. Available: https://dl.acm.org/doi/10.1145/3533708 (visited on 08/22/2023).

[34] L. Visuña, J. Garcia-Blas, and J. Carretero, “Novel Deep Learning-Based Technique for Tuberculosis Bacilli Detection in Sputum Microscopy,” in *International Conference on Interactive Collaborative Robotics*, Springer, 2023, pp. 269–279.

[35] European Commission, *Ethics guidelines for trustworthy AI | Shaping Europe's digital future*, en, Apr. 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (visited on 08/23/2023).

[36] European Parliament, *EU AI Act: First regulation on artificial intelligence | News | European Parliament*, en, Aug. 2023. [Online]. Available: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (visited on 08/23/2023).

[37] S. M. Park, K. Georgiev, A. Ilyas, G. Leclerc, and A. Madry, *TRAK: Attributing Model Behavior at Scale*, arXiv:2303.14186 [cs, stat], Apr. 2023. DOI: 10.48550/arXiv.2303.14186. [Online]. Available: http://arxiv.org/abs/2303.14186 (visited on 05/18/2023).

[38] D. Holzmüller, V. Zaverkin, J. Kästner, and I. Steinwart, *A Framework and Benchmark for Deep Batch Active Learning for Regression*, arXiv:2203.09410 [cs, stat], Aug. 2023. DOI: 10.48550/arXiv.2203.09410. [Online]. Available: http://arxiv.org/abs/2203.09410 (visited on 08/23/2023).

[39] Z. Liu *et al.*, *Influence Selection for Active Learning*, arXiv:2108.09331 [cs], Aug. 2021. DOI: 10.48550/arXiv.2108.09331. [Online]. Available: http://arxiv.org/abs/2108.09331 (visited on 08/23/2023).

[40] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, *Meta-Learning in Neural Networks: A Survey*, arXiv:2004.05439 [cs, stat], Nov. 2020. DOI: 10.48550/arXiv.2004.05439. [Online]. Available: http://arxiv.org/abs/2004.05439 (visited on 08/20/2023).

[41] J. Schmidhuber, "Evolutionary principles in self-referential learning. on learning now to learn: The meta-meta-meta...-hook," Diploma Thesis, Technische Universitat Munchen, Germany, Mar. 1987. [Online]. Available: http://www.idsia.ch/~juergen/diploma.html.

[42] C. Finn, P. Abbeel, and S. Levine, *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks*, arXiv:1703.03400 [cs], Jul. 2017. DOI: 10.48550/arXiv.1703.03400. [Online]. Available: http://arxiv.org/abs/1703.03400 (visited on 08/20/2023).

[43] Z. Chen, Y. Deng, Y. Wu, Q. Gu, and Y. Li, *Towards Understanding Mixture of Experts in Deep Learning*, arXiv:2208.02813 [cs, stat], Aug. 2022. DOI: 10.48550/arXiv.2208.02813. [Online]. Available: http://arxiv.org/abs/2208.02813 (visited on 08/22/2023).

[44] C. Hwang *et al.*, *Tutel: Adaptive Mixture-of-Experts at Scale*, arXiv:2206.03382 [cs], Jun. 2023. [Online]. Available: http://arxiv.org/abs/2206.03382 (visited on 07/14/2023).

[45] B. Mustafa, C. Riquelme, J. Puigcerver, R. Jenatton, and N. Houlsby, *Multimodal Contrastive Learning with LIMoE: The Language-Image Mixture of Experts*, arXiv:2206.02770 [cs], Jun. 2022. DOI: 10.48550/arXiv.2206.02770. [Online]. Available: http://arxiv.org/abs/2206.02770 (visited on 08/22/2023).

[46] N. Shazeer *et al.*, *Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer*, en, Jan. 2017. [Online]. Available: https://arxiv.org/abs/1701.06538v1 (visited on 08/23/2023).

[47] K. Minoura, K. Abe, H. Nam, H. Nishikawa, and T. Shimamura, *ScMM: Mixture-of-Experts multimodal deep generative model for single-cell multiomics data analysis*, en, Pages: 2021.02.18.431907 Section: New Results, Feb. 2021. DOI: 10.1101/2021.02.18.431907. [Online]. Available: https://www.biorxiv.org/content/10.1101/2021.02.18.431907v1 (visited on 08/22/2023).

APPENDIX A