Coupling Neural Networks Between Clusters for Better Personalized Care

Mathias Kraus
FAU Erlangen-Nürnberg
mathias.kraus@fau.de

Nico Hambauer
FAU Erlangen-Nürnberg
nico.hambauer@fau.de

Kristina Müller
FAU Erlangen-Nürnberg
annakristina.mueller@fau.de

Pavlina Kröckel
FAU Erlangen-Nürnberg
pavlina.kroeckel@fau.de

Nalika Ulapane
La Trobe University
nalika.ulapane@gmail.com

Arno De Caigny
IESEG School of Management
a.de-caigny@ieseg.fr

Koen De Bock
Audencia Business School
kdebock@audencia.com

Kristof Coussement
IESEG School of Management
k.coussement@ieseg.fr

Abstract

Personalized healthcare powered by machine learning (ML) is at the forefront of modern medicine, promising to optimize treatment outcomes, reduce adverse effects, and improve patient satisfaction. However, simple ML models generally lack the complexity to accurately model individual characteristics, while powerful ML models require large amounts of data, which are often unavailable in the healthcare domain. We address this problem with cluster-level personalization. In this method, similar patients are grouped into clusters and a local ML model is trained for each cluster. Since the amount of patient data to train ML models naturally decreases for each cluster, we introduce a novel objective function called "coupling" that allows information to be shared between clusters, so that smaller clusters can also benefit from information from larger clusters, thereby improving patient outcome prediction. Our method provides a compromise between a single global model for all patients and completely independent local cluster models. We show that coupling leads to statistically significant improvements on a simulated and a real-world dataset in the context of diabetes.

1. Introduction

Personalized healthcare has gained significant attention in recent years as an approach to tailor medical treatments based on individual patient characteristics, genetic information, and lifestyle habits [1, 2]. This approach aims to optimize treatment outcomes and minimize adverse effects, ultimately leading to a higher quality of care and better patient satisfaction. Personalized healthcare is becoming increasingly feasible due to advances in genomics, data analytics, but also due to the development of technologies such as smart wearables which give more personalized information about patients [3, 4]. Besides these developments, training of ML models that perform accurately on a personalized level is difficult. One of the most pressing issues is the lack of large-scale and diverse data for training these ML models [5]. Collecting such data is often challenging due to privacy concerns, limited access to patient records, or incompatible data sources [6].

In general, there are two ways to develop ML models to make predictions on a personalized level. First, very powerful and complex ML models can be trained which can accurately model individual's characteristics [7, 8]. To achieve this, these models require large amounts of diverse patient data, a condition that's often difficult to fulfill. The second approach is to train ML models for each individual [9]. While this approach naturally leads to a high degree of personalization, it is difficult to train a robust ML model due to only few and homogeneous datapoints available. Transferring information from one patient to another could potentially mitigate this challenge. However, the question remains from which patients the information should be transferred.

In response to these challenges, a promising approach emerges in the form of cluster-level personalization [10, 11]. Instead of building ML models that exclusively rely on individual or the whole set of datapoints, this approach groups similar datapoints or, in the context of medicine, patients into clusters and creates a shared machine learning model for each cluster [10]. To this end, we name these models (local) cluster models. These models allow us to leverage data from multiple patients who share similar characteristics, thereby substantially enhancing the available data for model training. It also enables early reliable predictions for new patients based on their cluster membership.

However, this approach must find a balance between between-cluster heterogeneity and within-cluster

URI: https://hdl.handle.net/10125/106821
978-0-9981331-7-1

homogeneity. By doing so, we leverage the shared characteristics within each cluster while still recognizing and utilizing the distinct variations between different clusters [10]. This balance ensures that the unique characteristics of each patient are considered, thereby moving towards personalized care while addressing the limitations of individual-level data scarcity. As such, we attempt to answer the following overarching research question: How can personalized healthcare be realized through cluster-level learning?

This paper aims to make three key contributions to the field of personalized healthcare and machine learning.

1. We propose coupling, a novel objective function for training neural networks on varying clusters.[1] By incorporating the coupling mechanism to train local cluster models, we manage to exploit the heterogeneity between clusters for precise model training, which results in more robust and accurate models.

2. We use both a simulated dataset and a real-world dataset in the context of diabetes to verify our approach. We study in depth, how clustering can lead to improved predictive performance, both for independent and coupled cluster models.

3. We discuss the generality of our approach by indicating that it can be applied to any neural network architecture, from simple feed-forward networks to more complex structures like recurrent neural networks, convolutional neural networks, or transformers.

We find that coupling leads to statistically significant improvements ranging between 0.6% and 7.0% lower mean squared error. The advantage of coupling was also evident in the cluster-specific performance: the coupled model consistently performed better in smaller clusters. We also find that coupling can lead to the behavior of a single global model, or independent cluster models. We show that we can achieve both behaviors by setting the regularization strength for our coupling approach.

Our work is relevant for healthcare organizations in multiple ways. First, in the realm of chronic disease management, such as diabetes, cardiovascular diseases, and mental health disorders, our approach can aid in developing precise treatment plans. Second, healthcare providers could effectively predict disease progression based on a patient's cluster, enabling proactive management of chronic conditions. This could lead to fewer hospital admissions, decreased healthcare costs, and improved patient quality of life. Third, pharmaceutical companies could use this method to conduct more efficient clinical trials. By identifying clusters of patients who are likely to respond similarly to a drug, researchers could better target their trials, reducing the time and costs associated with bringing a new drug to market. Fourth, healthcare policymakers can use this method for population health management. By identifying clusters of individuals who share certain characteristics and health outcomes, policymakers can tailor interventions at a community level, potentially reducing disparities in health outcomes among different population groups. Finally, our approach is also suitable for collaborative data usage by transferring information between naturally existing patient clusters, e.g., data from different hospitals. Since no raw data is exchanged between the clusters, coupling helps to address the problem of data scarcity and homogeneity while ensuring data privacy.

The remainder of the paper is structured as follows. Section 2 gives an overview over related literature from the fields of personalized healthcare, federated learning, and cluster-and-predict ML models. Section 3 describes our novel algorithm mathematically. Section 4 describes the way in which we simulated our dataset for glucose management. Section 5 introduces our evaluation strategy and Section 6 compares our approach against other, global, or fully decentralized methods, and Section 7 discusses our findings.

2. Background

This research lies at the intersection of personalized healthcare, federated learning, and cluster-and-predict models. In the following we introduce the related work in the respective fields.

2.1. Personalized Healthcare

Personalized healthcare is an approach that takes into consideration individual differences when providing medical treatment [1, 2]. The ultimate aim is to offer the right treatment to the right patient at the right time, based on a detailed understanding of their unique characteristics, such as genetic makeup, lifestyle, and environment [12]. By tailoring treatments to individuals, personalized healthcare aims to improve the effectiveness of care.

A major component of personalized healthcare is predictive analytics [5]. Predictive analytics utilizes machine learning models to predict patient outcomes based on historical data. These approaches are becoming increasingly feasible due to advances in genomics, data analytics, and other technologies such

[1] All code will be made available at `https://github.com/MathiasKraus/coupling-neural-networks`.

as smart wearables which give more personalized information about patients [3]. However, despite its potential, personalized healthcare is still in its infant stages. One key challenge is the need for large and diverse datasets to train accurate predictive models. Moreover, a significant variability exists between patients, adding another layer of complexity which a single, global machine learning model often does not capture.

2.2. Federated Learning

Federated learning, a distributed machine learning approach, offers a promising solution to the data scarcity problem [13, 14]. By enabling the training of machine learning models on decentralized data sources, federated learning allows for the creation of more robust models while maintaining patient privacy. Instead of sharing raw patient data, model updates are shared among participating devices, which are then aggregated to improve the global model [15]. This method not only alleviates privacy concerns but also enables the utilization of diverse and large-scale data, which is crucial for the development of accurate personalized treatments [16].

Our coupling algorithm can be seen as an adaption of the traditional federated learning algorithm [17]. Each cluster trains a local ML model on its centralized data corpus and then shares the model parameters with the other clusters, which incorporate the obtained parameters into the training of their local model. In this way, each cluster benefits from the knowledge gained in the other clusters while no raw data is shared between clusters. In the field of federated learning, a related idea has been implemented using a global model that shares information with the local cluster models to increase robustness [18]. In contrast, we do not rely on a global model and directly share parameters between the clusters (decentralized approach).

2.3. Cluster-and-Predict Models

Cluster-and-predict ML models are a class of machine learning algorithms that combine the strengths of clustering techniques and predictive models [19, 11]. These models aim to identify and leverage the underlying structure in the data, segmenting it into distinct groups, or clusters, based on the similarity of feature values. After the data is partitioned, a predictive model is trained separately for each cluster, accounting for the unique characteristics and patterns specific to that group. This approach is particularly effective when the data exhibits significant heterogeneity or when there are interactions between the features that can be better captured by dividing the dataset into homogeneous subgroups. By training separate models for each cluster, cluster-and-predict ML models can improve the overall prediction performance, providing a more accurate and nuanced understanding of the data [10, 19].

One notable example of a cluster-and-predict ML model is the logit leaf model, which combines decision trees and logistic regression [10]. In this approach, a decision tree is first used to split the feature space into clusters based on specific criteria or thresholds. The tree structure enables the model to handle complex interactions between features and identify distinct regions in the feature space. Once the clusters are formed, a logistic regression model is trained for each leaf node in the decision tree. By fitting a separate logistic regression model for each cluster, the logit leaf model can capture the unique patterns and relationships within each group, leading to more accurate predictions than a single global model [10].

Although these models have shown strong predictive performance, they have a major drawback of not fully exploiting the information provided by the dataset. This is because the feature space is split into disjoint clusters i.e. small clusters with very little data disregarding the information gained in other clusters.

As a remedy, we propose a novel approach which clusters the feature space, but allows information to pass between the clusters. We describe this approach in the following.

3. Method

This section presents a novel approach for combining clustering and prediction within a single framework, building upon the concept of cluster-and-predict ML models. Figure 1 illustrates our setting. Here, $n = 10$ patients are clustered into three groups based on their patient data. For each group, we want to train a cluster-level ML model. In prior work, each group trained an individual cluster-level ML model or a global ML model (ignoring the groups). In contrast, our approach allows to share information through a novel objective function designed to encourage similar patient behavior across clusters. We call this approach "coupling" since the local cluster models are not entirely independent but rather share information. This design resembles federated learning, where clusters act as independent data sources that exchange knowledge gained from their data.

3.1. Clustering

The clustering phase can utilize various clustering algorithms or even expert knowledge. In our case, we

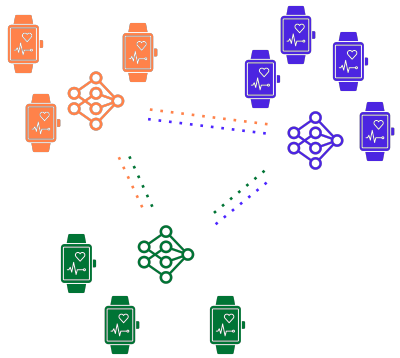

Figure 1. Illustration of our approach with three clusters of patients, where each cluster has their own cluster-level ML model that is connected to share information with other models.

opt for the widely-used k-means clustering algorithm due to its simplicity and effectiveness in partitioning the data into meaningful groups [20]. By dividing the data into clusters, we can identify groups of patients with similar behavior, laying the foundation for personalized predictive models in the next phase [21]. Note that clustering algorithms require access to the entire data space, thus privacy-preserving methods would require different clustering based on, for example, expert knowledge.

3.2. Prediction

For the prediction phase, we employ a specialized neural network architecture. Our key contribution lies in proposing coupling, a new objective function that seeks to capture similar patient behavior across clusters.

Mathematically, for patient information $X_1, \ldots, X_m$ for m clusters that can be used to predict the corresponding target values $y_1, \ldots, y_m$, we initialize m feed-forward neural networks $f_1, \ldots, f_m$ which include parameters $\theta_1, \ldots, \theta_m$ (also known as weights). Thus, a prediction within cluster j can be computed through $f_j(X_j, \theta_j)$. In case of a regression task, one commonly optimizes the parameters such, that the mean squared error between the prediction and the target values is minimized, i.e.,

$$\min_{\theta_j} \mathcal{L}_j = \min_{\theta_j} \Big(f_j(X_j, \theta_j), y_j \Big)^2, \tag{1}$$

where $\mathcal{L}_j$ is also known as the loss for cluster j.

For all clusters, we can define the optimization problem

$$\min_{\theta_1, \ldots, \theta_m} \sum_{j=1}^{m} \mathcal{L}_j, \tag{2}$$

which sums up the losses for all clusters. Note that θ_j can only influence the objective $\mathcal{L}_j$ and no other term, thus θ_j and θ_i are completely independent from each other for $i \neq j$ and no information is passed between the cluster models.

As a remedy, we propose the coupling of the cluster models. This allows the models to stay independent, i.e., between-cluster heterogeneity can be exploited to train precise cluster models, but information can be also shared. Figure 2 illustrates this idea.

Mathematically, we extend the optimization problem Equ. (2) through a regularization term,

$$\min_{\theta_1, \ldots, \theta_m} \sum_{j=1}^{m} \mathcal{L}_j + \alpha \underbrace{\sum_{i=1}^{m} \sum_{j=1}^{m} \big(\theta_i - \theta_j\big)^2}_{\mathcal{L}_{\text{couple}}}, \tag{3}$$

where $(\theta_i - \theta_j)^2$ measures the distance between corresponding parameters from cluster i and j and punishes high divergence. α denotes the regularization strength. Thus, for a large regularization term α, the cluster models tend to become very similar, as $(\theta_i - \theta_j)^2$ adds the highest weight to the overall objective. For a small regularization term, the cluster models tend to stay more independent from each other. As a result, our method offers several benefits:

Robustness. Clusters with few data points can still train reliable ML models, as information from other clusters contributes to stabilization.

Improved prediction performance. Each cluster trains its local cluster model, moving towards better prediction for personalized healthcare.

Interpretability. The coupling of cluster models enhances interpretability by allowing for an intuitive understanding of how different clusters relate to each other by observing shared parameters across models.

Cold start problem mitigation. Our method addresses the cold start issue, where model performance is generally poor until a sufficient amount of data is collected from a patient. In our case, the patient's cluster membership can already enables good model performance.

Generality. Our approach is general and can be applied to any neural network architecture. Although we use a simple feed-forward neural network in our case, more complex architectures such as recurrent neural networks, convolutional neural networks, and transformers can be employed directly.

In conclusion, our method extends the cluster-and-predict ML models by incorporating a coupling mechanism that facilitates information sharing among local cluster models. Our approach promotes

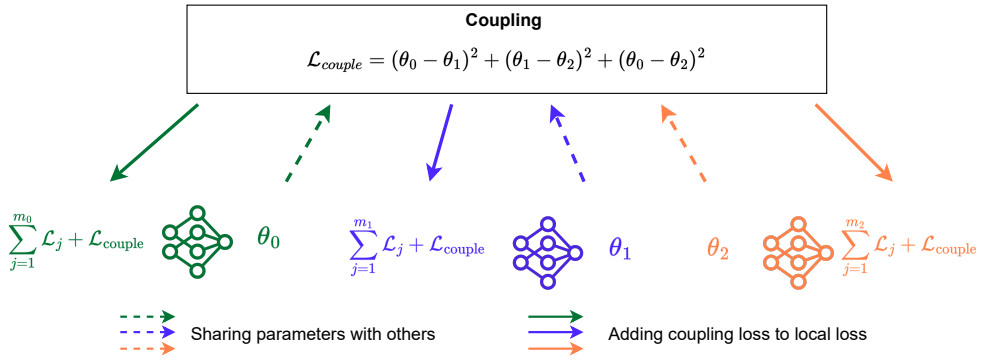

Figure 2. Illustration of our coupling approach. Parameters θ_0, θ_1, θ_2 define the local cluster models which are then shared with others to compute the coupling loss which is then added to the local cluster loss. By doing this, information can be shared between the individual models.

robust, personalized predictions, addresses the cold start problem, and offers a general solution applicable to various neural network architectures.

4. Data Simulation

In the context of diabetes management, effective insulin treatment is critical to maintaining glucose levels within a healthy range and preventing long-term complications [22, 4]. Personalized insulin treatment plans can lead to better glucose control, reduced risk of complications such as hypoglycemia and hyperglycemia, and improved overall quality of life for patients [23]. However, achieving this level of personalization requires continuous monitoring of glucose levels and the development of predictive models that can accurately estimate the effects of insulin on individual patients.

In the following, we describe the process to create simulated data for diabetes management in order to evaluate our proposed algorithm. The simulation resembles the case of diabetes management in which medical devices should accurately forecast the glucose level of patients (vector y) using six variables that represent patient information (matrix X). The method leverages clusters, which represent groups of patients with similar behavior within the group but different behavior between groups. In this simulation, the future glucose level should be influenced by the following six variables.

Insulin. Insulin is a hormone that helps regulate glucose levels in the blood by allowing cells to absorb and utilize glucose for energy. It represents the treatment which is generally optimized in order to control the glucose level.

Body-Mass-Index (BMI). BMI is a measure of body fat based on height and weight, which can impact how effectively the body processes glucose. Higher BMI values are often associated with increased insulin resistance, making it harder for the body to maintain healthy glucose levels.

Age. As people age, their body's ability to process glucose can decrease. Factors such as reduced muscle mass, changes in hormone levels, and decreased insulin sensitivity can contribute to age-related glucose fluctuations.

Gender. Men and women may have different responses to insulin and glucose metabolism due to hormonal differences and body composition.

Carbohydrates. Carbohydrates are a primary source of glucose. Consuming carbohydrates raises blood glucose levels.

Sports Activity. Engaging in sports or physical activity generally can lower blood sugar levels by increasing the uptake of glucose by the muscles, which use it for energy.

Algorithm 1 describes our data generation process. At the beginning of the procedure, we define the feature covariance matrix (Σ) to capture the correlations among the independent variables. We then establish the feature means (μ_j) and coefficients (β_j) for each cluster j. These means and coefficients represent the unique characteristics of each cluster, allowing us to differentiate between patient groups. Next, we assign a cluster (c_i) to each patient i. This assignment determines which group a patient belongs to and affects the relationship between the patient's information and their glucose level.

For each patient i in the dataset, we retrieve their cluster assignment j and the corresponding coefficients β_j and feature means μ_j. We then generate the patient's information X_i by sampling from a multivariate normal distribution with mean μ_j and covariance matrix Σ. This

process ensures that the patient's information correlates with the other features and is representative of the patient's cluster. Finally, we calculate the glucose level y_i for each patient by taking the dot product of their information X_i and the coefficients β_j and adding Gaussian noise.

Algorithm 1 Data Generation Process

procedure GENERATEDATA($n\_patients(np)$, $n\_clusters(nc)$)
  $\Sigma \leftarrow$ Define feature covariance matrix
  $\mu_j \leftarrow$ Define feature means, $j = 1, \ldots, nc$
  $\beta_j \leftarrow$ Define coefficients, $j = 1, \ldots, nc$
  $c_i \leftarrow$ Cluster assigment, $i = 1, \ldots, np$
  for each patient i in np **do**
    $j \leftarrow$ Cluster assignment of patient c_i
    $X_i \sim \mathcal{N}(\mu_j, \Sigma)$
    $y_i \leftarrow X_i \cdot \beta_j + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma)$
  return X, y

4.1. Sampling

In our simulation, we generate a total of 400 datapoints, with 200 datapoints allocated for training and another 200 for testing. This division allows us to evaluate the performance of our proposed algorithm on unseen data, which is crucial in assessing its generalizability and robustness.

We set the covariance matrix to

$$\Sigma = \begin{bmatrix} 1.0 & 0.3 & 0.1 & 0.0 & 0.4 & -0.3 \\ 0.3 & 1.0 & 0.2 & 0.1 & 0.2 & -0.5 \\ 0.1 & 0.2 & 1.0 & 0.0 & 0.1 & -0.2 \\ 0.0 & 0.1 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.4 & 0.2 & 0.1 & 0.0 & 1.0 & -0.1 \\ -0.3 & -0.5 & -0.2 & 0.0 & -0.1 & 1.0 \end{bmatrix}$$

and sample the feature means μ_j from $\mathcal{U}(-2, 2)$ and the coefficients β_j from $\mathcal{U}(0, 1.5)$. We sample the noise ϵ from a standard normal distribution $\mathcal{N}(0, 1)$.

In real-world scenarios, it is common to observe a skewed distribution of patients across different clusters [24]. Some clusters may contain a majority of patients, reflecting common patterns in patient behavior or characteristics. To simulate this phenomenon, we assign patients to clusters based on probabilities. Specifically, we assign a probability of 0.1 to cluster 0, 0.3 to cluster 1, and 0.6 to cluster 2. This probabilistic assignment means that the majority of patients will belong to cluster 2, mirroring the likely presence of large clusters in actual patient populations.

While we initially assume the existence of three clusters in our data, we do not restrict our analysis to this assumption. To further validate the effectiveness of our data generation process and the subsequent algorithm, we also employ cluster detection algorithms, specifically the k-means algorithm. This approach allows us to verify the cluster assignments and the number of clusters without making any prior assumptions.

4.2. Visualization

Figure 3 shows three clusters of patients generated with our data generation process. We applied Principal Component Analysis (PCA) to reduce the dimensionality of the patient information (matrix X) [25]. Upon visualizing the reduced dataset, we observed distinct clusters formed by the different patients based on their patient information. The PCA visualization confirms the effectiveness of our data generation process in creating a dataset that exhibits both intra-cluster similarities and inter-cluster differences.

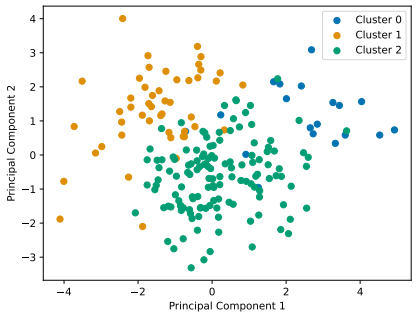

Figure 3. Visualization of three clusters generated with the data generation process. Dimensions are reduced using principal component analysis.

5. Evaluation

In this work, we primarily focus on linear models, which aligns well with the structure of our proposed coupling model. This choice is also reflective of the common practices within the field of diabetes management, where linear models have been frequently employed due to their interpretability and simplicity [26]. In the following, we describe the baseline models, our evaluation strategy, and the hyperparameter tuning.

5.1. Baseline Models

In order to evaluate the effectiveness of our proposed method, we compare it against two baseline models. The first baseline model is a global model that estimates the glucose level for the entire dataset, ignoring the presence of clusters in our data. This model serves as a benchmark for a traditional, non-personalized approach to glucose level prediction.

The second baseline model consists of local cluster models, which are trained completely independently from each other. This model represents a more

personalized approach to glucose level prediction, as it takes into account the unique characteristics of each cluster. However, it does not incorporate any mechanism for information sharing between clusters, which is a key feature of our proposed method.

5.2. Evaluation Strategy

Our evaluation strategy is divided into two stages. In the first stage, we evaluate the performance of the baseline models and our proposed coupled model under the assumption that the cluster assignments are known, i.e., relying on an "oracle". This stage allows us to assess the effectiveness of our method in a controlled setting, where the true cluster assignments are used to guide the model training process.

In the second stage, we relax this assumption and use the popular k-means clustering algorithm to predict the cluster assignments of the patients. This stage simulates a more realistic scenario, where the true cluster assignments are not known a priori and must be inferred from the data. By comparing the performance of our method under these two different settings, we can gain insights into its robustness and adaptability to varying conditions.

The performance of the models is evaluated using mean squared error (MSE) and mean absolute error (MAE) on the test set. We repeat the whole pipeline consisting of data generation, model training, model evaluation, 100 times for each model. Furthermore, we conduct a Wilcoxon signed-rank test to determine whether any observed differences in performance are statistically significant. Thereby, we compare corresponding models (e.g., local and coupled models for 2 clusters). This rigorous evaluation strategy ensures that our findings are robust and reliable, providing a solid basis for the potential application of our proposed method in real-world diabetes management scenarios.

5.3. Hyperparameter Tuning

Given our choice of model architecture, the main hyperparameter to tune is the regularization strength for the coupling, denoted as α in our objective function. To determine the optimal value for this parameter, we employ a grid search strategy over a range of potential values: [0.0001, 0.001, 0.01, 0.02, 0.05, 0.1, 0.5, 1.0, 10.0, 100.0, 1000.0]. The grid search allows us to explore a wide variety of potential regularization strengths, from very weak to very strong regularization, to find the balance that promotes shared information across clusters while allowing sufficient independence for individual cluster characteristics.

We train our neural networks for a total of 1000 epochs. Based on our observations, this number of epochs was sufficient for the models to fully converge, ensuring that we've achieved a stable model that is unlikely to improve further with additional training. Adam is used as the optimizer with a learning rate of 0.004 [27].

Lastly, for more efficient training, we initialize the weights of the cluster-level models with the weights from a globally trained linear regression model. This strategy leverages the knowledge from a preliminary global model to provide a good starting point for the individual cluster models, promoting faster convergence and improving overall computational efficiency. Again, it should be noted that similar to the clustering algorithm, global model initialization requires access to the entire data space, which is not a given in all conceivable use cases.

6. Results

In the following, we present the results of our model evaluations. In Table 1, we vary the cluster assignment of patients using an oracle, i.e., the correct assignment of each patient to the cluster that she was assigned to during the data generation process (line 5 in Algorithm 1), or a k-means clustering with different number of clusters. The statistical tests are performed between the coupled version and the corresponding fully independent, local version of the models.

We find that a global model performs worst across both MSE and MAE on our generated data. For perfect assignment, we find that coupling improves the MSE by 2.3 percent and the MAE by 0.9 percent. When using k-means for assigning patients to clusters, we find that coupling consistently outperforms the non-coupling models for both MSE and MAE. Improvements range between 0.6 and 4.0 percent.

We further observe that as the number of clusters increases from 2 to 4, the performance of the local model degrades gradually, with the MSE and MAE scores increasing. This trend indicates the challenges associated with creating effective local models when the true cluster structure is not accurately captured and the clusters become smaller and, thus, comprise less data to train robust ML models.

Conversely, the coupled model maintains relatively stable performance across different numbers of clusters, demonstrating its robustness to changes in cluster assignments.

6.1. Cluster Results Analysis

Table 2 provides the results for the MSE within each cluster. Notably, cluster 0, which is the smallest,

Model	Cluster Assignment	MSE	MAE
Global	—	1.414	0.944
Local	Oracle	1.147	0.853
Coupled	Oracle	1.121***	0.845***
Local	k-means(2)	1.313	0.910
Coupled	k-means(2)	1.305***	0.908***
Local	k-means(3)	1.329	0.914
Coupled	k-means(3)	1.303***	0.906***
Local	k-means(4)	1.383	0.931
Coupled	k-means(4)	1.330***	0.915***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1. Predictive performance of ML models for the simulated dataset.

Model	Cluster Assignment	Cluster 0 MSE	Cluster 1 MSE	Cluster 2 MSE
Global	—	2.028	1.546	1.239
Local	Oracle	1.709	1.130	1.063***
Coupled	Oracle	1.322***	1.133	1.091
Local	k-means(2)	1.967	1.344	1.181
Coupled	k-means(2)	1.910***	1.340	1.169**
Local	k-means(3)	1.933	1.336	1.223
Coupled	k-means(3)	1.822***	1.311***	1.179***
Local	k-means(4)	2.000	1.386	1.272
Coupled	k-means(4)	1.781***	1.328***	1.187***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 2. Predictive performance of ML models for individual clusters.

consistently shows significantly better results when coupling is applied, which indicates the effectiveness of the coupling approach in situations where the number of patients in a cluster is limited.

This is also in line with the finding that coupling becomes particularly important when the number of clusters is higher (e.g., k-means(3) or k-means(4)) which, again, naturally leads to less patients within each cluster. The coupled model consistently outperforms the local model under these conditions, showing significantly lower MSE values in all clusters.

6.2. Regularization Strength Analysis

Figure 4 presents the results for varying regularization strengths. The versatility of the coupled approach becomes apparent from these results. For instance, when a very high regularization value is used, the coupled model tends to behave similarly to the global model, with comparable MSE scores. This can be explained by the fact that the high regularization forces the parameters of the local models to adhere closely to each other, effectively producing a model that generalizes the data in a similar manner to the global approach.

On the other hand, when a very low regularization value is utilized, the coupled model shows behavior more akin to the local model, reflecting equivalent MSE scores. This is due to the low regularization allowing for the local cluster models to operate more independently from each other, which is a defining characteristic of the local approach.

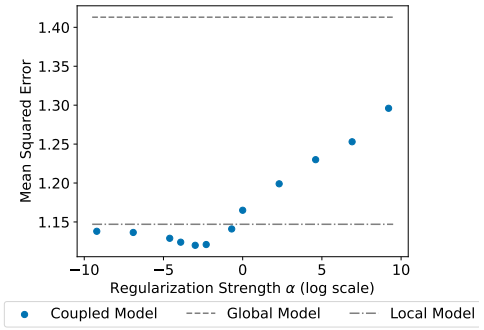

Figure 4. Mean squared error for varying regularization strengths. x-axis is in log scale.

6.3. Real-World Data

In the following, we use a real-world diabetes dataset [28]. The dataset includes ten baseline variables, age, sex, body mass index, mean blood pressure, and six blood serum measurements for 442 diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after baseline. We follow the same evaluation as before (see Section 5.2). Since we now do not know the true cluster assignment, we cannot evaluate the oracle models, but only models that use k-means clustering.

Table 3 shows our results. For two clusters, we find that independent cluster models slightly outperform coupled models. While the improvement is not significant for the MSE (p=0.80), it is on the MAE (p=0.03). However, for three and four clusters, the independent cluster models deteriorate in performance while the cluster models keep the performance or even improve slightly. We find the largest improvements for the MSE of 7.0 percent for four clusters. This observation is in line with our previous finding in that the benefit from coupling is especially pronounced for small clusters of data.

7. Discussion

We attempted to answer the overarching research question: How can personalized healthcare be realized through cluster-level learning? Our research provides valuable insights into personalized healthcare, demonstrating that cluster-level personalization,

Model	Cluster Assignment	MSE	MAE
Global	—	0.509	0.577
Local	k-means(2)	0.502	0.568**
Coupled	k-means(2)	0.503	0.571
Local	k-means(3)	0.515	0.574
Coupled	k-means(3)	0.502***	0.570***
Local	k-means(4)	0.538	0.585
Coupled	k-means(4)	0.503***	0.570***

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3. Predictive performance of ML models on our real-world dataset.

supported by our novel coupling function, can improve prediction outcomes for individual patients while benefiting from shared information across clusters. This finding bridges the gap between global and local machine learning models, offering an effective compromise that preserves data privacy while optimizing prediction outcomes.

Traditionally, most machine learning models in healthcare have been global [7, 8], presuming a single model can capture the inherent complexity and diversity of individual patients. These powerful models, though potentially able to handle this complexity, require large datasets and often lack transparency and robustness, limiting their suitability for many healthcare applications [29].

In contrast, local machine learning models, which are trained exclusively on data from individual patients, offer higher levels of personalization but may suffer from limited access to data, negatively affecting their performance [9]. In this regard, our coupling approach serves as an intermediary, allowing patient-specific models to benefit from shared information across clusters.

This approach also offers significant advantages from a federated learning perspective. Coupling can enhance the performance of local models trained on private data without the need to share sensitive patient data across different models [18]. In our method, only model parameters are shared, thus maintaining the privacy of the data within each cluster.

7.1. Limitations

While our research yields promising results, we want to discuss several limitations. First, our main study relies on simulated data. While this has allowed us to take a closer look into the model performance, it can not represent the complexity of real-world data. Second, our study primarily focuses on linear effects and the potential of coupling within a neural network context. Hence, the outcomes might differ when applied to other model architectures. Third, our research utilized k-means clustering to create patient clusters. While this method was effective for our needs, alternative clustering methods could yield different results and potentially improved predictive performance. Finally, our current method assumes all model parameters are coupled with equal strength. Although this approach showed efficacy in our study, it might not optimally capture the characteristics of individual patients in all healthcare domains.

7.2. Future Work

Looking forward, our research opens several exciting avenues. One of the compelling aspects of our research is its applicability across diverse data types and machine learning model architectures. While we have demonstrated the efficacy of our approach within the context of diabetes, the potential applications of our method extend beyond this specific area. The cluster-level personalization could potentially enhance predictive modeling in any healthcare domain or even in other sectors, such as marketing where personalization is crucial for decision-making and outcomes [10, 30].

On a methodological level, the coupling function could be extended to a variety of model architectures, including deeper neural networks and more complex structures like recurrent neural networks, convolutional neural networks, or transformers. These architectures might be more capable of handling complex and high-dimensional data. Additionally, future studies could investigate alternative clustering methods such as decision tree-based clustering [10] or hierarchical clustering. These methods might create more natural and meaningful patient clusters. Finally, future iterations of our method could consider a sparse coupling approach, where some parameters are kept entirely independent while others are tightly coupled.

7.3. Conclusion

In conclusion, our work presents a novel method of integrating individualized healthcare through cluster-level machine learning models. The proposed coupling function is an effective compromise between global and local models, improving patient outcomes prediction, enhancing robustness, and maintaining data privacy. We look forward to the impact and potential further advancements this research may bring to personalized healthcare and machine learning.

References

[1] M. G. Aspinall and R. G. Hamermesh, "Realizing the promise of personalized medicine," *Harvard business review*, vol. 85, no. 10, p. 108, 2007.

[2] M. A. Hamburg and F. S. Collins, "The path to personalized medicine," *New England Journal of Medicine*, vol. 363, no. 4, pp. 301–304, 2010. PMID: 20551152.

[3] A.-K. Witte and R. Zarnekow, "Transforming personal healthcare through technology-a systematic literature review of wearable sensors for medical application," in *Hawaii International Conference on System Sciences (HICSS)*, 2019.

[4] E. van Weenen, N. Banholzer, S. Föll, T. Zueger, F. Y. Fontana, K. Skroce, C. Hayes, M. Kraus, S. Feuerriegel, V. Lehmann, *et al.*, "Glycaemic patterns of male professional athletes with type 1 diabetes during exercise, recovery and sleep: Retrospective, observational study over an entire competitive season.," *Diabetes, Obesity & Metabolism*, 2023.

[5] V. Jagadeeswari, V. Subramaniyaswamy, R. Logesh, and V. Vijayakumar, "A study on medical internet of things and big data in personalized healthcare system," *Health information science and systems*, vol. 6, pp. 1–20, 2018.

[6] R. G. Fichman, R. Kohli, and R. Krishnan, "Editorial overview—the role of information systems in healthcare: current research and future trends," *Information systems research*, vol. 22, no. 3, pp. 419–428, 2011.

[7] Z. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *The New England journal of medicine*, vol. 375, no. 13, p. 1216, 2016.

[8] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 4–21, 2016.

[9] M. Maritsch, S. Föll, V. Lehmann, C. Bérubé, M. Kraus, S. Feuerriegel, T. Kowatsch, T. Züger, C. Stettler, E. Fleisch, *et al.*, "Towards wearable-based hypoglycemia detection and warning in diabetes," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–8, 2020.

[10] A. De Caigny, K. Coussement, and K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *European Journal of Operational Research*, vol. 269, no. 2, pp. 760–772, 2018.

[11] S. Brüggenjürgen, N. Schaaf, P. Kerschke, and M. F. Huber, "Mixture of decision trees for interpretable machine learning," *arXiv preprint arXiv:2211.14617*, 2022.

[12] E. A. Ashley, "Towards precision medicine," *Nature Reviews Genetics*, vol. 17, no. 9, pp. 507–522, 2016.

[13] B. Pfitzner, N. Steckhan, and B. Arnrich, "Federated learning in a medical context: A systematic literature review," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 2, pp. 1–31, 2021.

[14] I. Dayan, H. R. Roth, A. Zhong, A. Harouni, A. Gentili, A. Z. Abidin, A. Liu, A. B. Costa, B. J. Wood, C.-S. Tsai, *et al.*, "Federated learning for predicting clinical outcomes in patients with covid-19," *Nature medicine*, vol. 27, no. 10, pp. 1735–1743, 2021.

[15] Y. S. Can and C. Ersoy, "Privacy-preserving federated deep learning for wearable iot-based biomedical monitoring," *ACM Transactions on Internet Technology (TOIT)*, vol. 21, no. 1, pp. 1–17, 2021.

[16] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.

[17] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, "Federated learning of deep networks using model averaging," *ArXiv*, vol. abs/1602.05629, 2016.

[18] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.

[19] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *Machine learning*, vol. 59, pp. 161–205, 2005.

[20] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.

[21] N. M. Cohen, O. Schwartzman, R. Jaschek, A. Lifshitz, M. Hoichman, R. Balicer, L. I. Shlush, G. Barbash, and A. Tanay, "Personalized lab test models to quantify disease potentials in healthy individuals," *Nature medicine*, vol. 27, no. 9, pp. 1582–1591, 2021.

[22] R. Nimri, T. Battelino, L. M. Laffel, R. H. Slover, D. Schatz, S. A. Weinzimer, K. Dovc, T. Danne, and M. Phillip, "Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes," *Nature medicine*, vol. 26, no. 9, pp. 1380–1384, 2020.

[23] L. Bally, H. Thabit, S. Hartnell, E. Andereggen, Y. Ruan, M. E. Wilinska, M. L. Evans, M. M. Wertli, A. P. Coll, C. Stettler, *et al.*, "Closed-loop insulin delivery for glycemic control in noncritical care," *New England Journal of Medicine*, vol. 379, no. 6, pp. 547–556, 2018.

[24] D. Gartner and R. Padman, "Mathematical modelling and cluster analysis in healthcare analytics-the case of length of stay management.," in *International Conference on Information Systems*, 2016.

[25] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[26] V. Lehmann, T. Zueger, M. Maritsch, M. Kraus, C. Albrecht, C. Bérubé, S. Feuerriegel, F. Wortmann, T. Kowatsch, N. Styger, *et al.*, "Machine learning for non-invasive sensing of hypoglycaemia while driving in people with diabetes," *Diabetes, Obesity and Metabolism*, 2023.

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[28] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407 – 499, 2004.

[29] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature machine intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[30] A. De Caigny, K. Coussement, K. W. De Bock, and S. Lessmann, "Incorporating textual information in customer churn prediction models based on a convolutional neural network," *International Journal of Forecasting*, vol. 36, no. 4, pp. 1563–1578, 2020.