Exceptional Model Mining Based Active Learning Method

Alexandru Rosca
a.rosca@student.tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Kalina Bakardzhieva
k.bakardzhieva@student.tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Luca Mainardi
l.mainardi@student.tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Luyang Xie
l.xie@student.tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

Vincent Hoogendam
v.p.hoogendam@student.tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

ABSTRACT

Active Learning has become a prominent approach in machine learning for improving model performance while minimizing the labeling effort. Traditional active learning techniques techniques, such as uncertainty sampling, might be prone to oversampling certain regions and thus introducing bias in the model. To address this issues, we propose a new sample selection method that integrates Exceptional Model Mining (EMM) into the process. EMM identifies subgroups within the data that exhibit significant deviations from the global model, indicating potentially informative samples. The subgroups' descriptions can be beneficial in explaining the unlabeled set and making informed decisions about new samples to label. Furthermore, by selecting samples from these subgroups, our approach creates a more diverse and representative training set, therefore leading to more efficient training. We evaluate the performance of our method against traditional classification and traditional active learning, demonstrating that it has good classification power when the labeling budget is low.

KEYWORDS

Active Learning, Exceptional Model Mining, Machine Learning, Sample Selection, Data Mining

ACM Reference Format:
Alexandru Rosca, Kalina Bakardzhieva, Luca Mainardi, Luyang Xie, and Vincent Hoogendam. 2025. Exceptional Model Mining Based Active Learning Method. In *Research Topic in Data Mining, October, 2024 TUE, Eindhoven.* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The rapid expansion of machine learning is transforming various industries, driving advancements in areas such as healthcare, finance, and autonomous systems [3]. Machine Learning works best when vast amounts of data are available, however, this is not always the case. The challenge lies not only in gathering the volume of data but also in how effectively it is utilized to train models.

KDD '25, August, 2025, Eindhoven

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

Active Learning is a machine learning approach used when the costs to obtain a true label for the training samples is high. This could be because the labeling could be time-consuming, expensive, requires large amounts of manual effort or expertise. It aims to reduce the number of labeled samples needed by iteratively selecting the most informative data points for labeling. Some samples are more informative for the model than others. When presenting a model with a training sample that is very similar to what it has already seen, it is unlikely to learn from it, hence, by actively selecting which samples to label, active learning reduces the amount of samples needed.[9]

Active learning faces challenges that limit its efficiency. Traditional methods, such as uncertainty sampling and Query-By-Committee, often lead to redundant sample selection, increased computational costs, and bias in sample selection. These limitations can result in inefficient use of the labeling budget without providing enough informative samples for learning [5]. Furthermore, existing active learning methods offer little to no interpretability as to why a certain sample get selected[6].

To address the bias and redundancy problem, this paper introduces a novel sample selection technique based on Exceptional Model Mining (EMM). EMM identifies subgroups within a dataset where the data deviates significantly from the global model. We use these subgroups to gain insight into the characteristics of the unlabeled space and to indicate potentially informative samples. Thus, we are able to make more informed decisions about the selected samples and reduce the bias in the training set by selecting samples from different subgroups.

The contributions of this paper are twofold: (1) we use EMM to identify informative subgroups within the unlabeled space of the dataset, allowing to increase the interpretability of the selected samples, and (2) we develop a complete EMM-based active learning algorithm that uses these subgroups to create a diverse and unbiased training set.

2 BACKGROUND

In this section, we provide an overview of the two main concepts used in our research: *Active Learning* and *Exceptional Model Mining* (EMM).

2.1 Active Learning

Active learning is a machine learning approach used when obtaining labels for the training data is expensive. This can be due to variety of reasons, for example, if an opinion of expert is required which requires manual investigating of each sample. Another reason could be that a difficult or risky procedure is needed, for example, performing a biopsy on a patient. The main idea of active learning is to reduce the number of labeled samples needed by selecting the most informative samples to be labeled [9].

2.1.1 Fundamental Strategies. Several strategies have been developed for active learning. The first one is *Uncertainty Sampling*: The model selects instances for which it has the least confidence in its predictions [2]. For example, in classification tasks, this could be instances where the predicted probabilities are closest to the decision boundary. The second strategy is *Query-By-Committee (QBC)* which maintains a committee of models trained on the current labeled data [10]. The instances that produce the most disagreement among the committee members are selected for labeling. Query by committee is only possible when several models are trained that produce different uncertainty scores. These models can be for example the trees in a random forest. The Uncertainty sampling method only requires one model to be trained. Because of this uncertainty sampling is simpler to implement and faster to use so we proceed this report with uncertainty based sampling.

2.1.2 Challenges. Active learning has been applied in various domains such as natural language processing, computer vision, and bioinformatics [3]. However, it suffers from some problems. One of them is the cold start problem - the initial model is trained on a very small datasets which can cause unreliable confidence scores. Since these are used for the selection of future samples for labeling this might cause further problems[8]. Another problem is the model bias. The active learning may oversample certain regions where the uncertainty is high and thus create bias in the model [7]. Finally, active learning could be computationally expensive since it requires retraining and using the model on the whole dataset at each iteration.

The problems we target with in this paper is the sampling bias. We propose a method which is still based on the computed confidence scores but samples from different regions of the unlabeled space thus reducing the sampling bias. The method is based on Exceptional Model Mining which is explained in the next section.

2.2 Exceptional Model Mining (EMM)

Exceptional Model Mining is a sub-field of data mining focused on identifying subgroups within a dataset where the subgroup deviates significantly in some way from the global dataset. [4].

Formally, consider a dataset D consisting a set of descriptive attributes A with state space $\mathcal{A}$ and target attributes T with state space $\mathcal{T}$. The goal of EMM is to find subgroups $S \subseteq D$, defined by a selector s, such that the model M_S built on S differs significantly from the global model M_D built on the entire dataset D, according to some quality measure $\phi(M_S, M_D)$. Hence the search objective becomes to identify selectors s and search in the space that maximizes the quality measure: $s^* = \arg\max_s \phi\ (M_S, M_D)$.

3 SOLUTION APPROACH

In active learning, the aim is to select the most informative unlabeled samples to label, thereby improving model performance efficiently [9]. Traditional methods often rely on uncertainty measures to select individual instances.

We propose using EMM to identify subgroups within the unlabeled data where the model exhibits exceptional behavior, indicating regions where the data is different form the rest of the data by some measure. The key idea here is that the model would then select samples that are representative of more different subgroups and get a more refined, well rounded model. Furthermore, EMM outputs why some sample is in a certain subgroup based on the descriptors of the subgroup the sample is in. This makes EMM inherently interpretable[4].

3.1 Mathematical Formulation

Let us formalize the problem:

3.1.1 Data Setup. Let $D = L \cup U$ be the dataset, where:
$L = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n_L}$ is the initial labeled set, selected randomly for the first model training
$U = \{\mathbf{x}_i\}_{i=1}^{n_U}$ is the unlabeled set.
$\mathbf{x}_i \in \mathbb{R}^d$ represents the descriptive attributes for $i \in 1..8$ $y_i \in \mathcal{Y}$ represents the label with $\mathcal{Y} \in [0, 1]$
We have a classifier $M_D : \mathbb{R}^d \rightarrow \mathcal{Y}$ trained on L.

3.1.2 Defining the Target Attribute for U. For each unlabeled instance $\mathbf{x}_i \in U$, the classifier M_D trained on L provides a prediction $\hat{y}_i$ and a confidence score c_i where:

$$\hat{y}_i = M_D\ (\mathbf{x}_i)$$

$$c_i = \max_{y \in \mathcal{Y}} P_{M_D}\ (y \mid \mathbf{x}_i)$$

Here, $P_{M_D}\ (y \mid \mathbf{x}_i)$ is the predicted probability of class y for instance $\mathbf{x}_i$.

3.1.3 Target for EMM. The target attribute we define for the EMM step is the confidence score c_i over each unlabelled sample. Thus, we define the target attribute $t_i = c_i$ for each $\mathbf{x}_i \in U$. The set we perform EMM over is U' where:

$$U' = \{(\mathbf{x}_i, t_i)\}_{i=1}^{n_U}$$

3.1.4 Exceptional Subgroup Selection. Define a subgroup $S \subseteq U$ based on some selector function $s : \mathbb{R}^d \rightarrow 0, 1$, such that: $S = \{\mathbf{x}, \mathbf{t} \in U \mid s(\mathbf{x}) = 1\}$ Our goal is to find a selector s that defines a subgroup where the model M_S behaves exceptionally compared to its behavior on D.
To find such a selector function s, we shall use the BeamSearch algorithm. BeamSearch takes parameters w and d_{max}, which define the width and depth of the algorithm. Beam Search is a heuristic search algorithm that explores the search space by maintaining a fixed

number of the best candidate subgroups at each level, defined by the beam width w. The algorithm works as follows: it begins with the most general subgroup and iteratively expands it by adding new conditions to create more specific subgroups, up to a maximum depth d_{max}. Then at each iteration, the algorithm performs the following steps:
Expansion: for each subgroup, it refines the subgroup by adding one feature constraint at a time.
Evaluation: It calculates the quality of the subgroups using the quality measure ϕ we defined earlier.
Selection: Keep the top w subgroups based on their quality scores. These subgroups will be expanded in the next iteration. By focusing on the most promising subgroups, the BeamSearch algorithm allows us to efficiently explore the search space without considering every possible option, resulting in high-quality subgroups found in reasonable time.[4].

3.1.5 Quality Measures. We define three quality measures to quantify how exceptional the subgroup is compared to the global data.

The first measure, the *mean uncertainty deviation*, is mathematically expressed as:

$$\phi_1(S) = \mu_c S - \mu_c U = \frac{\sum_{i \in S} c_i}{|S|} - \frac{\sum_{i \in U} c_i}{|U|}$$

Where $|S|$ is the size of the subgroup and $|U|$ is that of the entire dataset. This formula calculates the mean deviation of the subgroup's confidence score from the overall dataset's uncertainty mean, highlighting regions within the data space that significantly differ in confidence levels.

The second measure, the *uncertainty z-score*, enhances this approach by considering the variability within the subgroup itself, and is defined as:

$$\phi_2(S) = \frac{\mu_c S - \mu_c U}{\sigma_S / \sqrt{|S|}}$$

where σ_S represents the standard deviation of uncertainties within the subgroup. This measure is particularly suitable for identifying subgroups whose uncertainties are not only different but statistically significantly different from the global mean, ideal for detecting outliers or anomalous subgroup behaviors.

Finally, the *Weighted Relative Accuracy Change (WRACC)* incorporates an entropic factor that lends weight to the relative size and diversity of the subgroup, formulated as:

$$\phi_3(S) = (\mu_c S - \mu_c U) \times \text{entropy}(S)$$

Entropy for a subgroup is calculated using:

$$\text{entropy}(S) = -\left(\frac{|S|}{|D|} \log_2 \frac{|S|}{|D|} + \frac{|S^c|}{|D|} \log_2 \frac{|S^c|}{|D|} \right)$$

where S^c represents the complement of the subgroup and $|D|$ is the total size of the dataset. This measure favors subgroups that are not only significant for their deviation from the norm but also have a balanced representation in terms of confidence scores, thereby promoting a comprehensive exploration of the feature space.

3.1.6 Optimization Problem. Our objective is to find the selector s that defines the subgroup S maximizing the quality measure ϕ. The optimal subgroup $s*$ would then be:

$$s* = \arg\max_s \phi(S)$$

In active learning often multiple samples are picked to be labeled at each iteration, this creates the risk of picking samples with similar features which creates redundant labeling costs[1]. Using the proposed approach, we can select the first n subgroups discovered by the EMM algorithm and select samples within each one of them. Thus, the selected samples to be labeled will represent different subgroups of the low confidence data, promoting diversity. Thus, the objective becomes to find the first n subgroups.

3.2 Integrating EMM into Active Learning

We propose the following algorithm for using EMM as a selection mechanism in active learning:

Algorithm 1 EMM-Based Active Learning

Require: Labeled training set $L = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n_L}$
Require: Unlabeled training set $U = \{\mathbf{x}_i\}_{i=1}^{n_U}$
Require: Test set $T = \{\mathbf{x}_i, y_i\}_{i=1}^{n_U}$

1: $M_D \leftarrow$ Train model on (L)
2: Accuracy $\leftarrow$ Evaluate(M_D, T)
3: $C = \{c_i =$ Confidence of M_D on $U | \forall \mathbf{x}_i \in U\}$
4: $S = \{s_i = \arg_{i \in top-10} \max_s \phi(S, C)\}$
5: $S^* \leftarrow$ Select n samples from each group in S
6: $L \leftarrow L \cup \{(\mathbf{x}, y) \mid \mathbf{x} \in S^*\}$
7: Repeat until labeling budget exhausted or sufficient Accuracy achieved

4 EXPERIMENTAL RESULTS

4.1 Qualitative Analysis of EMM on Pima Indians Diabetes Dataset

The experiments utilize the Pima Indians Diabetes Database, which consists of medical diagnostic data for predicting diabetes. In this subsection we perform a qualitative analysis on the performance of EMM on it.

The data is available publicly on https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database. All code used for setting-up and running the experiments can be found on our GitHub Repository: https://github.com/luca-mainardi/RTDM.

The dataset contains records for 768 female patients. The attributes included in the dataset capture a variety of physiological measurements relevant to diabetes diagnosis. The attributes are the following: the number of pregnancies each patient has had, the plasma glucose concentration (measured two hours after an oral glucose tolerance test), the diastolic blood pressure (millimeters of mercury), the triceps skin fold thickness (millimeters), the serum insulin levels (micro-units per milliliter), the Body Mass Index (BMI), a diabetes pedigree function, which estimates the patient's hereditary risk of diabetes and the patient's age. The target variable is a binary outcome indicating the presence (1) or absence (0) of diabetes. We choose this specific dataset because it contained only numerical values, had a good number of descriptive attributes and one target variable. Reducing complexity introduces by the dataset as much

as possible allows us to focus solely on the research question.

EMM can utilize different search algorithms for identifying the subgroups. For this experiment we use EMM Beam Search approach[4]. We want to achieve group descriptions which are both interpretable and expressive so we choose to have a bin depth of 3 for the beam search. This means each subgroup will be described by at most 3 attributes. Furthermore, we fix the bin width to 10, meaning the algorithm discovers 10 subgroups per iteration. To avoid groups of only a couple of elements, we restrict the group size to 5% of the whole dataset.

For the use of the discovered subgroups in selection of active learning samples, it is preferable that the subgroups have no overlap. In this experiment we implement a pruning mechanism that penalizes overlap within the groups, the algorithm used for this is an adaptation from the one used in this GitHub Repository: https://github.com/MathynS/emm. Essentially, we compute and maximize the coverage (proportion of the group that does not intersect another group).

Using different quality measures of the EMM results in different groups. Here, we perform a qualitative analysis to investigate how the quality measure impacts the resulting groups. The quality measures compared are mean uncertainty deviation, z-score deviation and weighted relative accuracy uncertainty. The different quality measures produce groups of different sizes.

	Description	Score	Size
1	$130 < \text{Insulin} \leq 744$ AND $115 < \text{Glucose} \leq 140$	0.26	48.5
2	$140 < \text{Glucose} \leq 18.57$ AND $3 < \text{Pregnancies} \leq 6$	0.23	37
3	$130 < \text{Insulin} \leq 744$ AND $0 < \text{Pregnancies} \leq 1$	0.25	31
	...		
	Total Size		425

Table 1: Description of first 3 groups using Mean Uncertainty Deviation

	Description	Score	Size
1	$130 < \text{Insulin} \leq 744$ AND $115 < \text{Glucose} \leq 140$	8.5.56	48.5
2	$130 < \text{Insulin} \leq 744$	8.74	147
3	$36 < \text{BMI} \leq 67$ AND $140 < \text{Glucose} \leq 18.57$	6.55	52
	...		
	Total Size		1045

Table 2: Description of first 3 groups using Z-score

The groups produced by mean uncertainty deviation are of similar size usually between 30 and 50. Furthermore, all of the groups are described by two attributes. The combined size of all ten groups is 425, which might include any overlapping samples regardless of the pruning technique used.

The groups produced by uncertainty z-score are usually bigger and of more heterogeneous sizes. Some of the groups are described by a single attribute. We also observe that comparing with the above method the first discovered group is the same but with uncertainty z-score the second group has very large overlap with the first with one identical identifier. The combined size is 1045 which signifies large overlaps considering the size of the full dataset. The groups produced by WRACC are also significantly bigger and the total size is 1330. Here all groups are described by a single attribute.

	Description	Score	Size
1	$130 < \text{Insulin} \leq 744$	0.15	147
2	$140 < \text{Glucose} \leq 18.57$	0.12	153
3	$41 < \text{Age} \leq 72$	0.08.5	136
	...		
	Total Size		1330

Table 3: Description of first 3 groups using WRACC

Due to this differences we can conclude that the mean uncertainty deviation produces the most informative groups with homogeneous sizes and with limited overlap on the Pima Indians Diabetes dataset.

4.2 Optimizing EMM for active learning

In this subsections we tune the EMM method in the context of active learning. We investigate different strategies for sample selection within the group, as well as how the number of samples within the group and the number of iterations impacts the results.

The active learning set up is as follows. Firstly, we withhold randomly selected 20% of the full dataset for testing. Then out of the remaining 615 samples, only 123 or 20% are initially labeled. Following the algorithm described in Section 3.2. we firstly train a model on these. The model we choose for the experiment is Naive Bayes since it is a popular, simple and fast classification model. The accuracy on the withheld test set is recorded. Then we use the model on the other 80% of the training set in order to identify the uncertainty score. The next step is to identify the subgroups, select new samples to label, retrain the model and record the accuracy on the test set. We do this for 10 iterations.

We consider two strategies for selecting samples for labeling within the group. The first is to select samples at random from each group. The second strategy is to select the most uncertain samples within each group. We perform the experiment with different number of samples labeled per iteration - 1 per group (or 10 in total), 2 per group (or 20 in total) and 5 (or 50 in total). Furthermore, we compare how the three quality measures investigated before perform.

Figure 1. and figure 2. show that the accuracy of the classifier increases significantly in the first couple of iterations even though only few samples have been added - 40 and 80 respectively in the first four iterations. These still constitutes less than 10% and less than 20% of the full dataset.

Figure 1. demonstrates that the uncertainty sampling technique performs better than the random sampling technique. However, there is no significant difference between the three quality measures. Figure 2. supports the same findings.

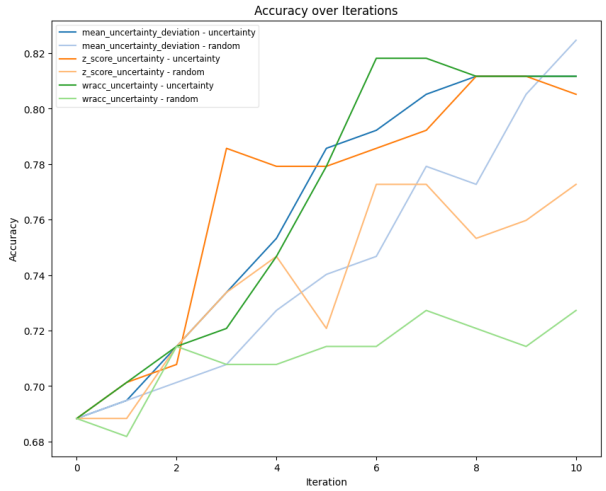

Figure 1: Comparison adding 1 sample per group per iteration

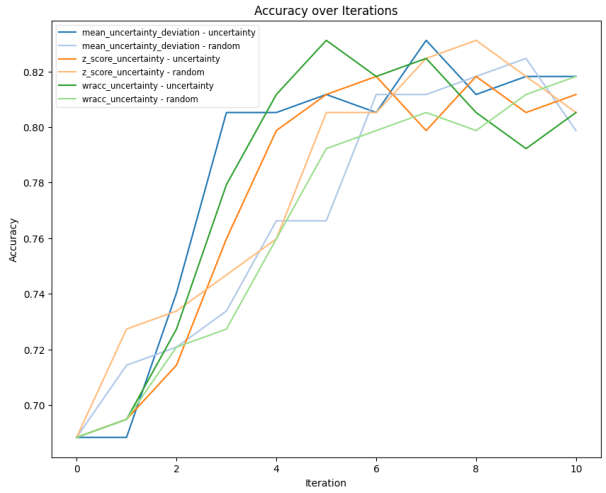

Figure 2: Comparison adding 2 samples per group per iteration

Figure 5. shows that when we have labeled sufficient number of samples the differences between the two sampling techniques become unimportant. Here we can also observe that when a big proportion of the dataset has been included in the training, the accuracy decreases which might be a sign that the model is overfitting. This is expected since we do not include any measures against overfitting in this study.

From this experiment we can conclude that the choice of the quality measure has little impact on the performance of the active learning classifier. However, the choice of sampling measure is important with the uncertainty sampling performing better. Once a sufficiently big proportion of the training set has been labeled the active learning parameters have no effect.

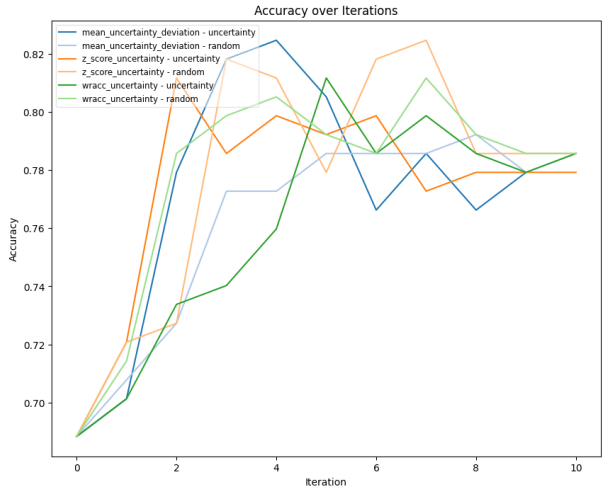

Figure 3: Comparison adding 5 samples per group per iteration

4.3 Comparing the EMM active learning to traditional active learning

In this subsection we compare the tuned EMM based active learning method to two baselines - traditional classification and uncertainty based active learning. We use the set up described in the above section. The test set is 20% of the data, the initial training set contains 20% of the remaining samples and the model used is Naive Bayes.

Since all quality measures performed similarly well on the quantitative analysis of the active learning, we proceed to use the mean uncertainty deviation as quality measure due to its good qualitative performance on the dataset. The sampling method within the groups is uncertainty based due to its definitively better performance over the random sampling.

For the baseline without active learning we simply randomly select and label the required number of samples from the unlabeled set. For the active learning method we utilize uncertainty sampling. Formally this mean that after calculating a confidence score c_i for each sample $\mathbf{x}_i \in U$, we identify the required number of instances with confidence score closest to 0.5:

$$\mathbf{X}_n^* = \underset{i \in \text{top-}n}{\arg\min} \left| c_i^{(t)} - 0.5 \right|$$

We obtain the true label $\mathbf{x}_i$ for $i \in X_n^*$ and update the labeled set: $L^{(t+1)} = L^{(t)} \cup \{(\mathbf{x}_{i^*}, y_{i^*})\}$ Update the unlabeled set: $U^{(t+1)} = U^{(t)} \setminus \{\mathbf{x}_{i^*}\}$ and retrain our classifier $M_D^* : \mathbb{R}^d \to \mathcal{Y}$ trained on L. Then we evaluate the performance of $M_D^{(t)}$ on a our test set. We iterate this process for T iterations. At each iteration we obtain a new classifier $M_D^{(t+1)}$ trained on the updated labeled set $L^{(t+1)}$.

We again perform the experiments with adding 1,2 or 5 samples for each of the 10 group per iteration. The results of the experiments are as presented below.

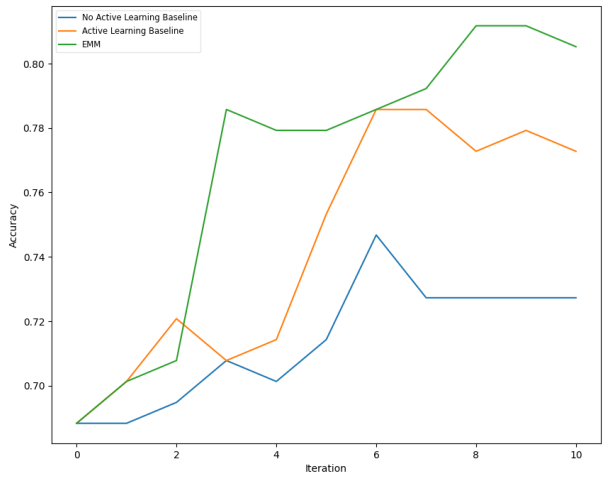

Figure 4: Comparison adding 1 sample per group per iteration

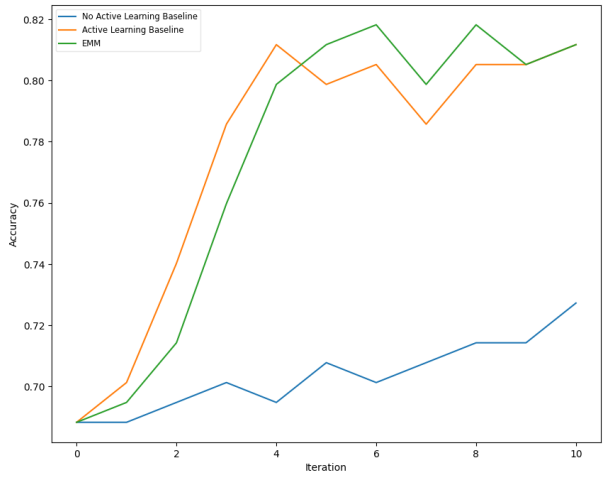

Figure 5: Comparison adding 2 samples per group per iteration

We observe that when adding one sample per iteration our method performs better than the active learning baseline and significantly better than the no active learning baseline. This supports our theory that using EMM is beneficial for selecting the set of most informative samples in active learning. When adding two samples per iteration, our method performed similarly to the active learning baseline and still significantly better than the no active learning baseline. Finally, we observe that after adding sufficiently large number of samples to the training set, the method for selecting them (randomly, uncertainty based or EMM based) does not make an impact, however, selecting such a large number of samples to be labeled will likely not be the case in most Active Learning settings. Furthermore, the accuracy going down might be due to overfitting.

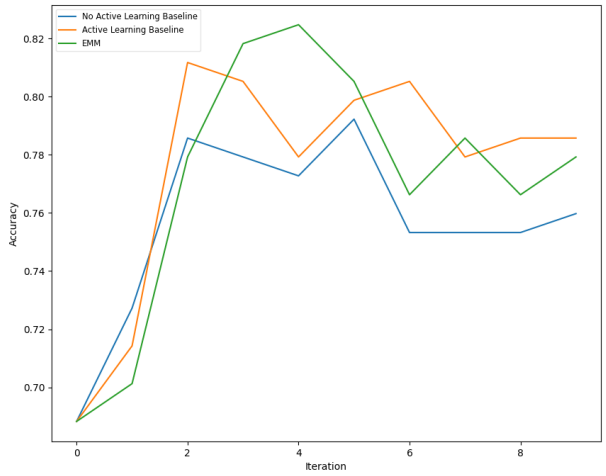

Figure 6: Comparison adding 5 samples per group per iteration

5 CONCLUSIONS

In this paper, we proposed a novel method for sample selection in the Active Learning process using Exceptional Model Mining (EMM). Traditional Active learning methods face, among other challenges, a high chance of introducing bias in the model by oversampling certain regions. By integrating EMM, our approach identifies subgroups within the unlabeled data whose uncertainty characteristics deviates from the rest of the data. By choosing samples from within each identified group, the training set includes more diverse and representative samples. Furthermore, the usage of EMM promotes interpretability of the active learning approach by giving a definition of the constraints of the subgroup.

Our experiment on the Pima Indians Diabetes dataset, show that EMM is capable of identifying descriptive groups within the unlabeled set which allows us to make a more informed selection of the samples for labeling. By labeling the most uncertain samples from the discovered subgroups we show that the EMM-based method performs better than selecting random samples for labeling and at least on par with uncertainty based active learning for the given classification task. The benefits of using the EMM based active learning are particularly visible when the labeling budget is small and only a small part of the dataset can be labeled per iteration.

Thus, the proposed EMM based active learning methods has been shown to have two main benefits. It provides descriptive information about the unlabeled samples in order to support informed selection process and it reduces the model bias by providing diverse samples for labeling.

Future Work

While our current work has demonstrated the potential of EMM-based sample selection, there are still several areas that remain open for future exploration.

- Future work could test other variations of EMM, including different search algorithms, exploring different quality measures, or different parameters.
- Our experiments focused on a single relatively small dataset. Applying this method to larger and more complex datasets, not limited to just numbers, would help evaluate the scalability and flexibility of the EMM based active learning approach.
- Currently, we have tested the EMM based active learning method only on classification tasks. However, it would be beneficial to examine how the method performs on other supervised tasks like regression.
- One limitation of active learning methods which we do not address is the computational cost. In our proposed method this is even higher since both the classification model and the EMM model need to be run at each iteration. Future work could focus on developing more efficient algorithms or heuristics for subgroup discovery and EMM evaluation.
- Currently we fixed or manually tuned the parameters in the EMM process in order to limit the complexity of the experiments. Future research could explore automatic tuning techniques based on the dataset's characteristics.

In conclusion, the proposed EMM based active learning method is a good alternative to traditional active learning creating more diverse and informed samples. Future work could maximize the potential of the method.

REFERENCES

[1] Jordan Ash, Chicheng Zhang, Akshay Krishnamurthy, John Langford, and Alekh Agarwal. 2020. Deep Batch Active Learning by Diverse, Uncertain Gradient Lower Bounds. (2020). https://doi.org/10.48550/arXiv.1906.03671 arXiv:1906.03671

[2] David Cohn, Les Atlas, and Richard Ladner. 1994. Improving generalization with active learning. *Machine Learning* 15, 2 (01 May 1994), 201–221. https://doi.org/10.1007/BF00993277

[3] Steven E. Dilsizian and Eliot L. Siegel. 2013. Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment. *Current Cardiology Reports* 16, 1 (Jan. 2013), 441. https://doi.org/10.1007/s11886-013-0441-8

[4] Wouter Duivesteijn, Ad J. Feelders, and Arno Knobbe. 2016. Exceptional Model Mining. *Data Mining and Knowledge Discovery* 30, 1 (01 Jan 2016), 47–98. https://doi.org/10.1007/s10618-015-0403-4

[5] Sheng-Jun Huang, Rong Jin, and Zhi-Hua Zhou. 2014. Active Learning by Querying Informative and Representative Examples. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 10 (2014), 1936–1949. https://doi.org/10.1109/TPAMI.2014.2307881

[6] Ishani Mondal and Debasis Ganguly. 2020. ALEX: Active Learning based Enhancement of a Model's Explainability. (2020). arXiv:2009.00859 https://arxiv.org/abs/2009.00859

[7] Chelsea Murray, James U. Allingham, Javier Antorán, and José Miguel Hernández-Lobato. 2021. Addressing Bias in Active Learning with Depth Uncertainty Networks... or Not. (2021). https://doi.org/2112.06926 arXiv:2112.06926

[8] Vishwesh Nath, Dong Yang, Holger R. Roth, and Daguang Xu. 2022. Warm Start Active Learning with Proxy Labels and Selection via Semi-supervised Fine-Tuning. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. Springer Nature Switzerland, Cham, 297–308.

[9] Burr Settles. 2009. *Active Learning Literature Survey.* Technical Report.

[10] H. S. Seung, M. Opper, and H. Sompolinsky. 1992. Query by committee. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (Pittsburgh, Pennsylvania, USA) *(COLT '92)*. Association for Computing Machinery, New York, NY, USA, 287–294. https://doi.org/10.1145/130385.130417